IRE Transactions

ON MILITARY ELECTRONICS

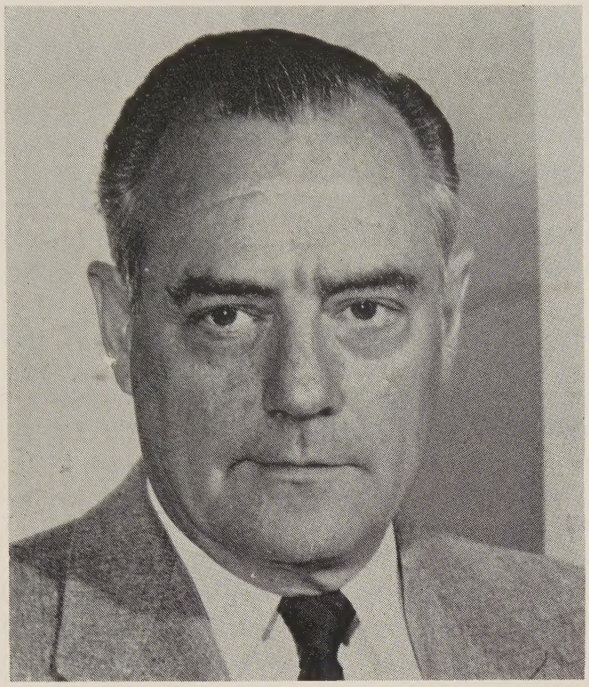
Volume MIL-5 APRIL, 1961 Number 2

ADVANCED RADAR TECHNIQUES

PUBLISHED BY THE
PROFESSIONAL GROUP ON MILITARY ELECTRONICS

IRE TRANSACTIONS® ON MILITARY ELECTRONICS

Published by The Institute of Radio Engineers, Inc., for the Professional Group on Military Electronics at 1 East 79 Street, New York 21, N. Y. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group, or its members. Individual copies of this issue, and all available back issues, except Vol. MIL-4, Nos. 2, 3, and 4, may be purchased at the following prices: IRE members (1 copy) $2.25, libraries and colleges $3.25, all others $4.50. Annual subscription rate: non-members $17.00; libraries and colleges $12.75.



James M. Bridges

James M. Bridges (SM'47–F'59) was born in Hancock, Me., on August 29, 1906. He received the B.S.E.E. degree from the University of Maine, Orono, in 1928.

From 1928 to 1942 he was a member of the Engineering Department staff of the New York Telephone Company. During the last seven years of this period he held the position of Radio Engineer and directed the establishment and installation of the New York Harbor Radio Telephone System and the Vehicular Emergency Radio Telephone System in New York City.

From 1942 to 1946, as an active officer in the Naval Reserve, he was associated with weapon control radar research and development in the Navy Bureau of Ordnance. He was separated from the service as a Commander. During this period he received commendation from Navy Secretary Forrestal for contribution to the advancement of Naval fire control.

From 1946 to 1952, as a Civil Service engineer, he held the position of Head of Radar Research and Development for the Navy Bureau of Ordnance, Washington, D.C., and in 1952 was appointed Chief Engineer for Electronics for that Bureau. Since May, 1955 he has been Director of Electronics in the Office of the Assistant Secretary of Defense (Engineering).

Mr. Bridges is a member of the American Ordnance Association and the Armed Forces Communications and Electronics Association, and is Chairman of the Advisory Group on Reliability of Electronic Equipment.

A New Generation of Radar

JAMES M. BRIDGES, *Guest Editor*

IN TRACING the history of radar development we associate with its early conception—or "invention" —the names of relatively few scientists and engineers. Such men as Young, Taylor, Watson-Watt, Blair, Page, and Hyland, are on record and generally accepted as the "fathers" of this tremendously important area of military technology. But we find it hard to credit to specific individuals very many of the remarkable technological advances that have been made during radar's two decades of existence. As in most of the rapidly growing fields of technology, few new ideas are conceived and developed entirely by a single individual, or even by a single agency. Rather, the technological advancement of radar has in most cases resulted from determined attacks on the limiting technical problems by thousands of dedicated scientists and engineers working in many government, university, and industrial laboratories. The contribution of one man or group adds to the achievements of others, and the feedback of knowledge acquired multiplies the total capability.

Largely through this extensive interchange and interaction of knowledge a new concept can mature fully, be physically realized, and become operationally useful. In this process it is hoped that this issue of the IRE TRANSACTIONS ON MILITARY ELECTRONICS, dealing with advanced radar techniques, will act as an effective catalyst.

As radar enters its third decade, it is appropriate that we review and summarize some of the most recently evolving technology, which has kept radar in its position of vital military importance throughout the period of rapidly advancing aircraft and missile performance and even into the era of space penetrations. Such a summary, of course, provides a sense of gratification to the thousands of people who created and fostered these technological advances. More than that, it creates a point of reference from which we can look back on our accomplishments, confirming our knowledge and acquiring the courage and stimulation we need to recognize and accept new challenges to even greater progress.

Even the small sample of advanced radar techniques covered in this issue makes it clear that a new generation of radar has been born. Radar ranges are expressed no longer in thousands of yards but in thousands of miles. The measurement of range and angular position, for years the only information directly obtained by radar, is no longer adequate. We must now measure velocities and acceleration as well. Not long ago the accuracy of range and angular position was measured in hundreds of yards and significant fractions of a degree; it must now be measured in terms of feet and fractions of an angular mil. Shifting of a radar beam's position in space, which for most military applications was satisfactory in degrees per second, must now be in terms of degrees per millisecond—sometimes even in degrees per microsecond.

The technology has to do more than meet these stringent demands for radar performance: Not only must the radar determine with great precision where the target is and where it is going, but it must describe the target for us. Is it round, flat, skinny or fat? Is it rotating or stabilized in space? The radar of the future must be able to secure all this information when the target is incased in an ion plasma, possibly operating in a medium disturbed by a nuclear explosion, or in the presence of several different kinds of radar countermeasures.

These are the kind of challenges that have faced the scientists and engineers who have been developing radar technology over the past few years, demanding that they pursue new techniques and develop new and exciting components. These requirements have brought together in a single radar system such sophisticated and complex techniques and components as pulse compression, maser or parametric amplifiers, pulse Doppler, digital data processors and superpower microwave generators. They have forced the development and successful application of electronically-scanned radar beams, or phased arrays.

During the 1950's the application of this fast-moving radar technology strikingly resulted in scores of different radar types for many varied military uses. These range all the way from a radar that a man can carry on his back and use to locate personnel moving in darkness or concealed in woods, to ballistic-missile-detecting radars whose antennas equal the size of three football fields standing on edge, end to end.

Our scientists and engineers have been eminently successful in supplying the techniques, components, and radar systems to satisfy these difficult and diverse operational requirements. At the same time, unfortunately, they have—perhaps necessarily—brought about a tremendous increase in complexity and systems cost. It is not uncommon, although still shocking, to see a proposal for a single radar costing a significant part of a billion dollars and calling for tens—even hundreds—of thousands of active electronic components.

I am convinced that, of all the challenges facing the future radar designer, the most important, and perhaps the most difficult, is a significant reduction in these two rapidly skyrocketing factors—cost and complexity. It is vital to continued success in the military application of radar that we do this. We are ably meeting the fantastic requirements for advanced radar technology in this age of ballistic missiles and space vehicles. But these successes will be of little value if we cannot afford

to buy as many of the new radars as we need, or if we cannot keep them operating after we have bought them. In the words of Dr. Eugene Fubini, "It is not entirely a question of whether or not a concept is feasible—the question is will it stay 'feased'?"

It appears that radar, in the foreseeable future, will continue to be a vital and critical component of our military capability. The radar designer will encounter demands for increasingly greater performance and new applications. He must develop and exploit new portions of the frequency spectrum between microwave and infrared wavelengths. Radarlike devices operating in the optical-wavelength region appear important and are perhaps not too far away. For space radars, drastically reduced size, weight and power consumption, coupled with extremely long life and high reliability, will be demanded. And there will be requests for radars with average radiated powers in the microwave (or millimeter) region of several megawatts. Based on these and other future requirements, problems will confront the radar designer that are even more difficult than those he has solved during the past decade.

As the creative technical resources available to carry out this work will not be unlimited, it is most important that they be used effectively—more effectively than ever before. We must develop new techniques and components and prove their feasibility in advance of systems development rather than to commit systems engineering with its tremendous demands on technical resources to unproven techniques or "unavailable" components.

As Guest Editor, and on behalf of the officers and members of PGMIL, I thank the contributors to this issue of the TRANSACTIONS, as well as the more than 100 persons who submitted excellent papers that could not be published here because of space limitations. We hope that another issue on radar techniques, with a broader coverage of the area, can be published soon.

The Future of Radar*

JOHN S. BURGESS†, SENIOR MEMBER, IRE

Summary—In this paper an attempt is made to forecast the direction in which the radar state-of-the-art will aim. The paper will begin with a description of a few of the radars of World War II, to set the stage for a comparison and evaluation of the various techniques that have been added.

Since World War II, various components and techniques developments have been designed for our more modern-day radar, and a tremendous increase has been realized in capability for handling smaller, faster and more distant targets. Particular emphasis is placed on the problems of high resolution, discrimination and pattern recognition.

The effect of our entry into the space age on the design of radar equipments will be discussed. These new radars must cope with problems which are at least an order of magnitude greater than the air-breathing threat in all of its aspects.

The conclusion made is that the phased-array type of radar offers the only long-range solution to the complex problems faced today by the radar engineer. Its combination of flexibility, limitless power, high-frequency capability, electronic scanning, etc., makes it the only logical choice for the future.

IT IS advantageous to sit back occasionally and examine from where one has come and to where one is going in any field of endeavor. When viewed on a day-to-day basis, progress may seem interminably slow but, when considered over the span of many years, the integrating effect of time and effort becomes very apparent. Two decades have passed since the perpetration of the infamous raid on Pearl Harbor at which time a new but untried radar, the AN/SCR-270, displayed blips of that impending disaster. It is the intent of this article to review in retrospect the tremendous advances made in the radar state-of-the-art during the past twenty years and to attempt to project into the future the directions along which we may find outselves traveling. Although examples are limited to ground-based equipments, the developments described apply equally well to other types of radars. The descriptions are general in nature, since detailed reviews are contained elsewhere in this issue and in the literature.

First, let us review some of the earliest radars. By "earliest" I do not mean to go back quite as far as the radars depicted by the artists in the Bomac series of advertisements, or even as far back as the experimental equipments of Sir Watson-Watt. Rather, I refer to radars such as the AN/SCR-270, developed by the Signal Corps Laboratories and produced by Westinghouse, which was one of our first World War II operational radars. Fig. 1 shows it in site in the Pacific area, ready to give warning of aircraft within a radius of 100 miles. It operated in the 100-Mc region with 100 kw of peak-power output. Accuracy was limited to about ±4 miles in range and to ±4 degrees in azimuth. It formed the eyes and ears of the earliest electronic command and control systems. Its purpose was to give warning of an impending attack of enemy aircraft. Such warning was relayed through the fairly extensive telephone and radio networks that existed throughout the Pacific area in conjunction with the "sky watch" program which the allies had set up throughout the South Sea Islands. The AN/SCR-270 was quite a successful radar for its time, and commanders relied upon it to a very great extent. For example, at one time the Fifth Air Force Headquarters was located at Brisbane, Australia, near a very important supply depot in Townsville to the north. Perhaps one of the earliest and most significant successes of the AN/SCR-270 was the warning it gave of an eighty-aircraft attack coming in to bomb the supply depot. The radar operator for the AN/SCR-270 spotted the incoming raid on the outer fringes of the PPI, one hundred and fifty miles out. He relayed this information to the fighter control center from which point air-

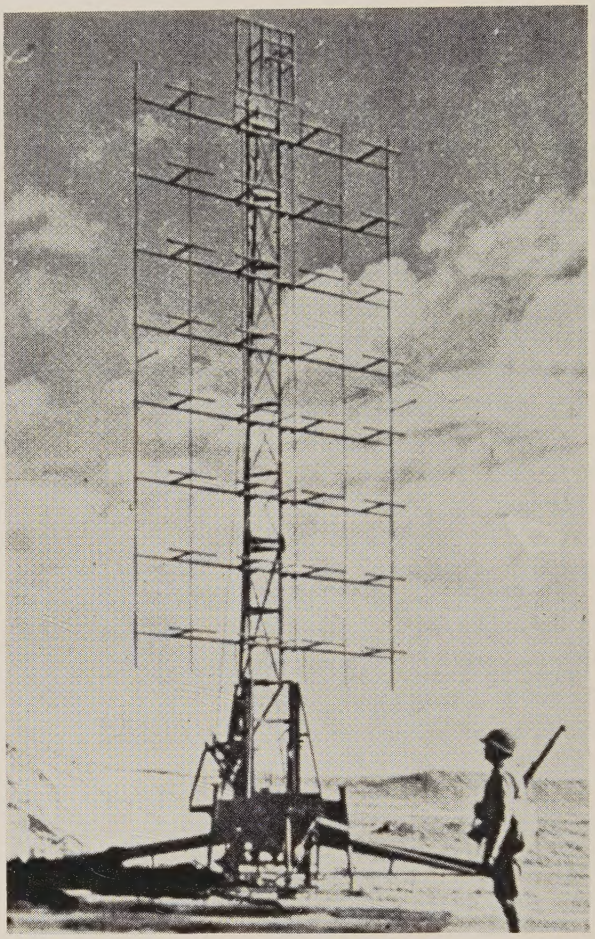


Fig. 1—AN/SCR-270.

* Received by the PGMIL, January 17, 1961.
† Rome Air Dev. Ctr., Griffiss AFB, N. Y.

craft were dispatched, and the enemy was engaged quite a distance from the supply depot. As a result, not one bomb was dropped on the depot nor did one enemy plane return from that engagement. In fact, this might be representative of the earliest air defense system wherein the radar operator read off range and azimuth from his display tube and gave this information to a plotter sitting at his side; the plotter then calculated the position of the target and relayed the data by either phone or radio to the fighter control headquarters from which point aircraft were dispatched immediately to engage the enemy.

Another group of early radars was the AN/TPS series. Shown in Fig. 2 is the AN/TPS-3 lightweight, transportable, search-and-tracking radar. This set operated in the 600-Mc region with a peak power of 200 kw, and exhibited a range of 100 miles on a bomber.

Transportability was an important factor in the design of radars for World War II. An extreme example of this is the AN/TPS-2 which could be carried by eight men, as shown in Fig. 3. Radar logistics presented a somewhat simpler problem than is prevalent today.

The tremendous strides that have been and are being made in radar technology are possible only through the close cooperation which exists between government laboratories as team leaders and the many educational and industrial research laboratories. The government scientist is responsible for translating present and future operational requirements into a meaningful and effective basic and applied research program. This involves not only the selection of a program from among the many ideas submitted, but also the justification and defense of this program in an endeavor to obtain a fair share of the appropriated funds. He must participate actively in an in-house research program so that he can better guide and understand the efforts of contractor laboratories. He must cull the fruits of this research program and translate them into operational equipments or systems, thus providing the military services with an improved capability. It is this interplay between governmental and industrial and educational laboratories which brought about the technological explosion of the last two decades.

Let us begin with the radar equation shown in (1):

$$R_{max} = K \sqrt[4]{\frac{P\tau G^2 \lambda^2 F^{1/3} \sigma}{N}} \qquad (1)$$

where

P = Peak power
τ = Pulse length
G = Gain of antenna
λ = Wavelength
F = Repetition rate
σ = Echo area
N = Noise figure.

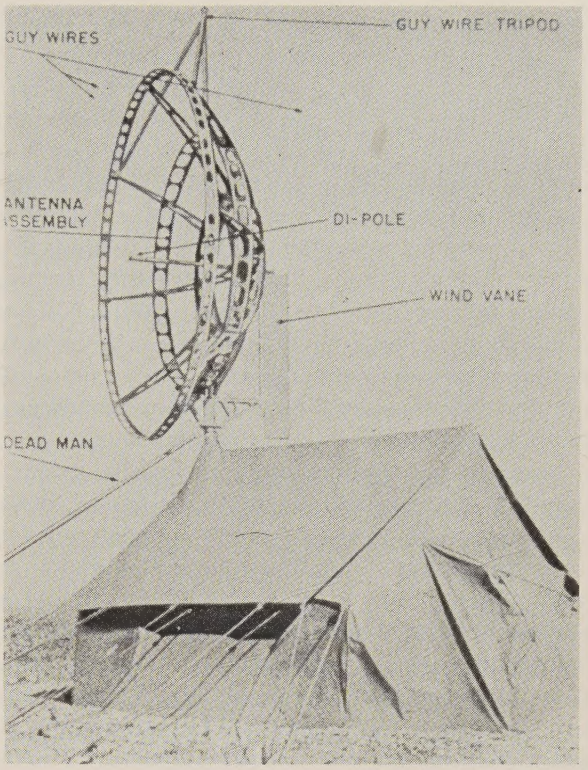


Fig. 2—AN/TPS-3.

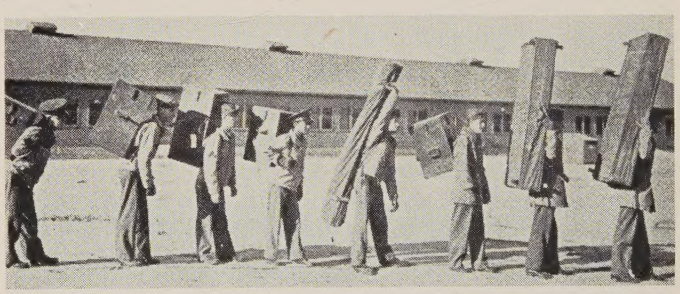


Fig. 3—Man-packed AN/TPS-2.

It, of course, has many forms, but this is one of the simpler ones which shows the factors available for increasing radar effectiveness. This equation is well known and is shown here merely to remind the reader of the various parameters which can be developed in improving our radar performance. The major components of a radar system are represented by the factors given in this equation. Since many papers in this issue deal specifically with the various component developments, this subject will not be treated in great detail; rather, emphasis will be placed on where we are on some of the major components and on the advancements made in this regard during the past few years.

Consider the subject of power. The only known method of generating the necessary power for radar applications in the microwave frequency region is through microwave tubes. Power output P_o from these tubes is affected by the factors shown in (2):

$$P_o \propto \eta j_0^2 \lambda^2 D \qquad (2)$$

where

η = Efficiency
j_0 = Current density of cathode
λ = Operating Wavelength
D = Dissipation factor.

The dissipation factor and efficiency are closely related but require completely different research approaches. As better materials and cooling techniques are developed, more heat can be dissipated for the same level of efficiency; and, as techniques are developed for increasing electronic conversion efficiency, more useful power can be obtained. Both of these research approaches hold promise of yielding higher-output powers. The importance of efficiency is apparent when one realizes that, for each unit of power dissipated, one unit of useful power is available at 50 per cent efficiency; whereas four units of useful power are available at 80 per cent efficiency and nine are available at 90 per cent.

As the wavelength of operation is reduced, the size of the physical structure decreases, complicating the generation of high power by reducing the size of the cathode and by reducing available area for dissipation. Consequently, considerable effort is being applied to cathode materials and to techniques for side-stepping the limitation in size. For example, RCA's new series of super-power tubes uses the approach of a large number of circuits operating in parallel.

In addition to the factors shown in (2), the design of a tube is strongly affected by desired gain, bandwidth of operation, tunability, average-vs-peak power required, etc. It is interesting to note that as yet there are only two basic principles available to the designers of microwave tubes; *i.e.*, the linear-beam tube and the crossed-field tube.

In the linear-beam interaction, as shown in Fig. 4, it can be assumed that by some means the electrons have been injected into the tube from a gun at some source located to the left of the picture. These electrons are bunched and, as they pass through the output cavity of a microwave tube, they contain a certain amount of kinetic energy which has been derived from the gun potentials. If the tube is so designed that the electrons, as they pass through the gap, are slowed down by an electric field in the gap, the electrons lose energy by conversion of kinetic energy to the RF field. When the RF field is 180 degrees out of phase, the kinetic energy of the electrons is increased; however, then, because of the bunching effect, no electrons are in the gap. By the time the next bunch of electrons appears, the field has reversed itself and is in phase to extract energy from it. The actual bunching effect is accomplished by velocity modulation of the electrons as they leave the gun area. A small RF field, superimposed on the relatively large dc voltage, produces the velocity modulation.

The other type of electronic interaction which occurs is in the crossed-field tube, shown in Fig. 5, wherein the electron is traveling with a velocity $v = E/B$ perpendicular to the lines of the magnetic field and perpendicular to the lines of the electric field. Due to this motion, the force on the electron is well known. It is the Lorentz force which will give it some curvature in the field. In this case, if the impressed RF field is in the proper phase, the electron gradually is moved toward the anode of the tube. Thus, the electron is losing its potential energy to this RF field. Inherently, this is a more efficient type of tube because of its capability to convert all of the potential energy of the electron into RF energy. The electron, of course, still retains some kinetic energy of motion across the field, although techniques have been developed for including some conversion of this kinetic energy to render even higher efficiencies.

These are the two available techniques for the design of tubes, and they have been used by tube engineers for many types of tube developments. Combinations and variations of these principles have resulted in the availability of a wide variety of tube types. In World War II, only two types of tubes were available—the space charge tubes such as triodes and tetrodes, and the brand-new magnetron. The trend in the modern radar, for several reasons, is to utilize the master-oscillator-power amplifier type of transmitter, primarily because it provides a great deal more flexibility for control of the

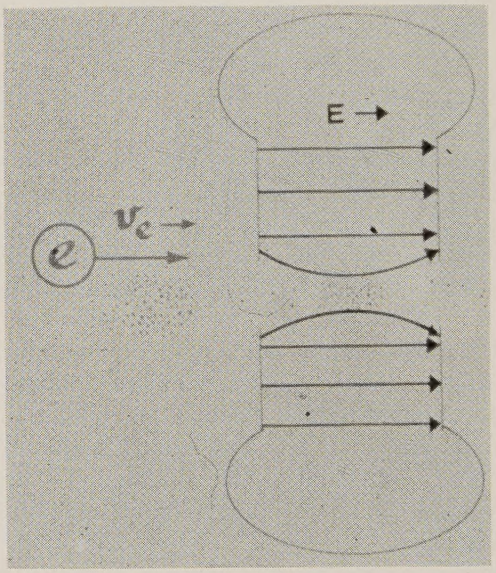


Fig. 4—Linear-beam interaction. Conversion from kinetic energy.

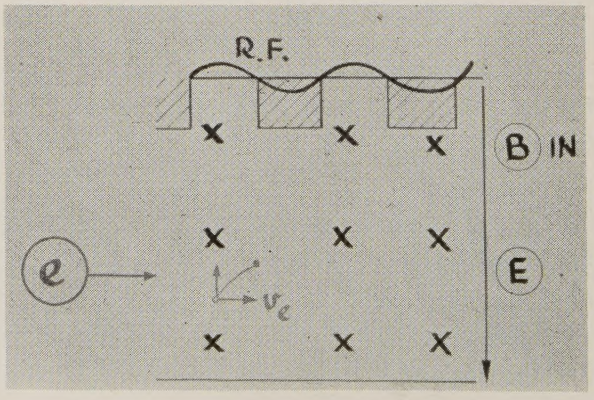


Fig. 5—Conversion from potential energy.

pulse shapes, coding, etc. Available peak-power levels have increased about an order of magnitude in each decade—from 100 kw in 1940 to 1 Mw in 1950 and to 10 Mw in 1960.

Going to the other extreme from the high-transmitter powers, the state-of-the-art in noise figures will be discussed next. During World War II, the noise figure of the radar set, in spite of its low-frequency operation, was around 15 to 25 db, and normally either crystal or triode circuits were employed. We now have added a fairly long list of techniques for achieving considerably lower-noise figures. Some of these are shown below:

	Noise Figures	
	State-of-Art	Future
World War II	15–25 db	
Crystals	8	5.5
TWT	6.5	1.5
Masers	<1 db	<1 db
Parametric amplifier	2 db	0.5 db
Tunnel diode	3.4 db	2 db
Triodes	9 db	9 db
Frequency—1250 Mc/Sec		

The crystal noise figure, as well as the triode, has been decreased by a considerable amount of engineering. In addition, we have traveling-wave tubes, masers, parametric amplifiers, tunnel diodes, and other types of circuits which are being devised continually in an effort to lower the noise figure of the radar set. The latest issues of the IRE Proceedings are filled with articles on these newer approaches. It is not too difficult to show that a 3-db increase in the radar range is much cheaper if obtained by lower-noise figures rather than by corresponding increase in power.

Two areas where particularly significant accomplishments have been realized during the past several years are the area of antennas and the area of receivers or signal processing. Let us first discuss signal processing. Eq. (3) shows essentially what initiated the requirement for advances in signal processing:

$$P_a = F\tau P \tag{3}$$

where

P_a = Average power.

The radar designer has the following dilemma. He knows that his range will increase as a function of average power proportional to the peak power, pulse width and the PRF; so, in order to increase his range, he must increase the average power by increasing any one or more of these three factors. However, the range resolution is inversely proportional to the pulse width. That is, better resolution is attained with smaller pulse widths. Consequently, increasing the average power by increasing the pulse width would decrease ability to resolve targets in range. In the case of the PRF, to increase the range the power pulses must be allowed to go out further and a longer period of time is required for the echo to return. Consequently, to increase the maximum unambiguous range the PRF must be decreased; so, increasing the average power by increasing the PRF works against an increase in range. The only recourse then is peak power. However, increasing the peak power above the 10 to 20 Mw of today introduces many technical problems of power breakdown, high voltage, X rays and so on. Therefore, the radar designer faces a true dilemma in trying to increase his average power by any one of these three methods.

A technique which attempts to overcome this dilemma and which has gained increasing attention during the past few years is pulse compression. Fig. 6 shows that an impulse or very narrow pulse is passed through a phase-distortion filter, the characteristics of which are such that the amplitude, as well as the phase, may vary with frequency. The result is that a stretched or elongated pulse appears at the output of the filter. This fairly long pulse then is amplified to a fairly high power and is transmitted as a long pulse. Consequently, an increase in the amount of average power or energy per pulse can be transmitted at relatively low peak powers.

Fig. 7 shows that this stretched pulse, as it returns from the target, is passed through a matched filter which is the inverse of the filter used for the transmitted pulse. In this case, the characteristic of the filter is such that the phase dependence is the negative of the original filter and the amplitude function is the same. The net result is a recovery of the original narrow pulse. Some residue or sidelobes of this pulse results, in view of the fact that the bandwidth of the filter cannot be matched perfectly to the shape of the pulse. At any rate, the end result is the high-resolution capability of the very narrow pulse. At the same time, a large amount of average

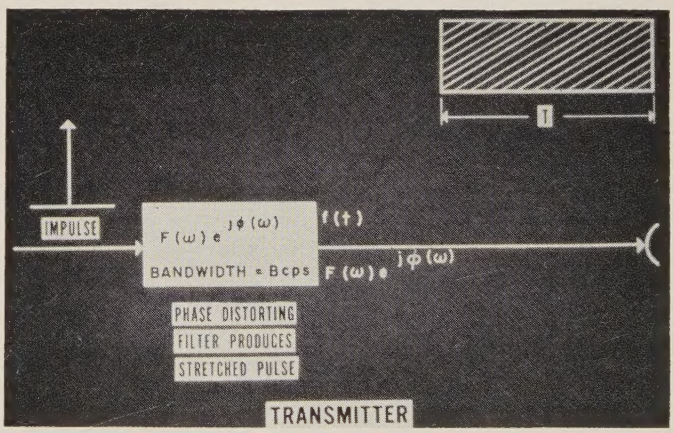


Fig. 6—Pulse-compression fundamentals.

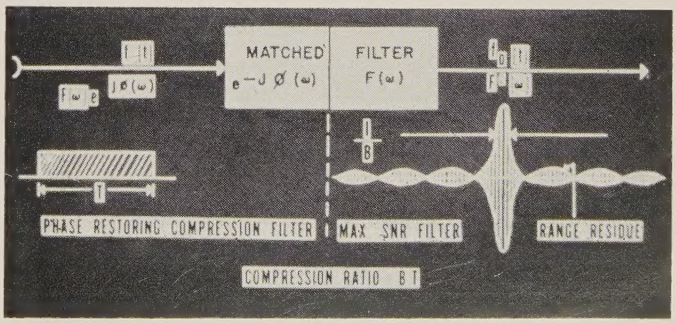


Fig. 7—Pulse-compression fundamentals. 3b receiver.

power has been transmitted. In this case, however, range ambiguities are introduced, in that range residues remain within the pulse. This effectively results in the same problem as that encountered with sidelobes in an antenna in that, if in range one large target and one small target are to be resolved within the same beamwidth, it is possible that the range residue response may be as great as that of the main narrow pulse.

Pulse-compression ratios in the order of a thousand have been achieved and proposed techniques allow a compression ratio of several thousand. Notice that the design of such a filter provides considerable flexibility. While variations in the amplitude function normally are not encountered, they can be. The phase function can be a linear FM, an alternating of some positive to negative phase values, a step function, or anything else which is limited only by the ingenuity of the engineer.

Another approach to the problem of achieving more energy on target is the work being done on phased arrays and modifications thereto. The basic principle involves feeding power to a large number of individual radiators with complete control of phase for each element. The phases can be prescribed so as to form a beam in any direction in space. This negates the real high-power problem of transmission lines, power generation of tubes, and so on, since a large number of the smaller power sources is used and the results are added in space to form a very-high-power beam. Another advantage of this approach is that, with electronic steering of this beam, it is not necessary to scan all of space continuously. By programming the beam to known tracks, a very efficient track-while-scan operation can be used. Another advantage of the phased array is that it can be flush-mounted and therefore given a fair degree of hardening. By combining phased arrays with Luneberg lenses or with frequency scanning of one sort or another, a great degree of flexibility can be designed into this type of radar.

The pincushion technique is another recent suggestion for a method of concentrating a large amount of power in a small beam. By this means, the wide aperture of a low-frequency radar is used at very high frequencies, to the extent that a large number of very narrow beams covers the same space as that covered by an equivalent single beam of a low frequency. In addition to placing more power on a target, this gives the added advantage of much higher resolution, both angular and velocity.

Also involved in the resolution problem is an attempt to reduce the effects of sidelobes in antenna characteristics. Hughes Aircraft Corporation recently completed an analysis of the possibility of modulating in time the aperture of a linear array or reflecting antenna. Such a modulation of phase and amplitude of the aperture illumination, as well as aperture size, has been considered, and it has been shown that a reduction in sidelobes in the order of 10 db can result. It may be possible to optimize such methods of modulation to reduce to an absolute minimum the sidelobes of a given antenna over a reasonably wide band.

Terminology such as "superesolution," "synthetic spectrum," "synthetic aperture," and so on, has recently come into vogue. These terms are applied to attempts made to obtain more energy on target without loss in resolution or position accuracy, and consist of either playing with the type of pulse which is transmitted, playing with the method of beam formation, or tinkering with the received signal.

A typical group of radars, known as the "frequency diversity" radars, is very close to field use. Although the detailed characteristics of these radars cannot be described, pictures can be shown and comparison can be made with the World War II radars. Fig. 8 shows the AN/FPS-35 radar designed by Sperry Gyroscope Company. Notice the more permanent type of structure that is used for the air defense radar of today. The antenna structure requires a five-story building to house the necessary components that make up the system. Fig. 9 shows a radar which does not even look like a conventional type. This uses the corporate structure antenna in which the power is divided and subdivided into a large number of individual waveguide feeds. It is the AN/FPS-28, manufactured by Raytheon.

In the early part of this discussion, reference was made to the pincushion technique. Fig. 10 shows an artist's sketch of a radar which is being built by Raytheon under ARPA sponsorship. Notice that cassegrainian optics are used for this system and the transmitter will sit right up on the antenna pedestal, rotating with the antenna structure.

So far, the modern radar has dealt with the problem of detection of aircraft. With the space age upon us, the radar now has the additional task of detecting and tracking long-range missiles, satellites and other spacecraft. Note that the term "radar" is used. This is because, to date, no brainstorms provided other means of meeting our present-day requirements if noncooperative targets are to be included. For this function, it might very well be expected that the shape of a radar must differ considerably, and so it does. Fig. 11 shows an experimental detection and tracking radar site installed by the Rome Air Development Center on the Island of Trinidad. This radar can look at long range missiles fired from the Atlantic Missile Range downrange toward Ascension Island. It also is in a strategic position to observe most of the satellites which are being placed in orbit and, in fact, is one of the important inputs into the National Space Surveillance Control Center of the Hanscom Complex.

The size of the required structure is necessarily fixed with respect to the ground so that limited scanning can be accomplished, and the antenna is pointed in the general direction in which a target might be expected. A series of sheet beams are placed into space where they narrow-beam-scan on each sheet. By this means, a target which penetrates one sheet can be detected, and its time of arrival to the next sheet gives sufficient information to determine trajectories and velocities. The information from this warning radar which has a range

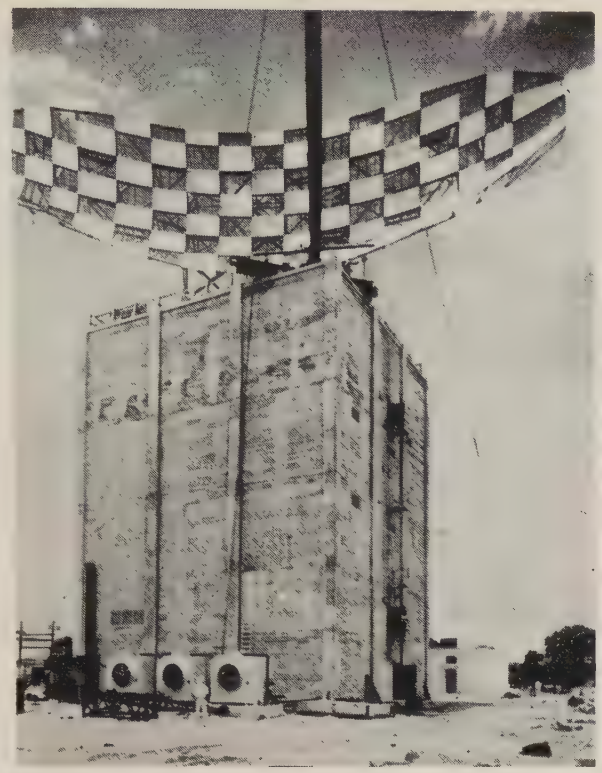


Fig. 8—AN/FPS-35.

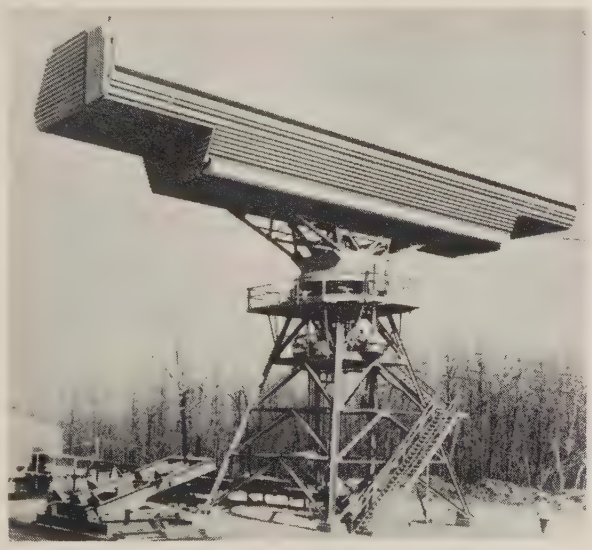


Fig. 9—AN/FPS-28.

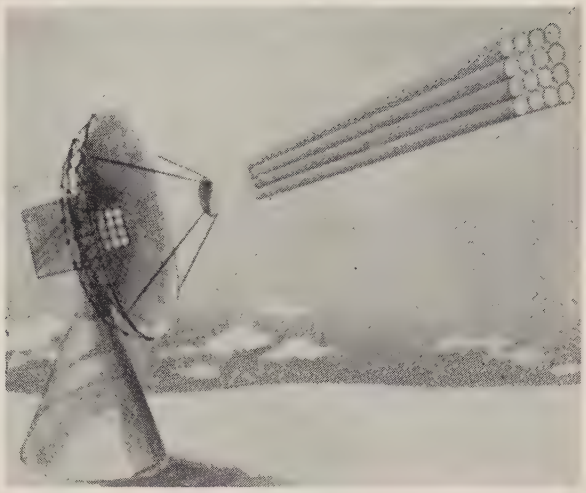


Fig. 10—Pincushion-tracking radar.

Fig. 11—RADC experimental radar at Trinidad.

of about 2000 miles is then used to slew the long range scanner onto the target, so that refined range and trajectory data can be obtained. In order to cover a fairly wide range of territory, a chain of such radars must be developed, as is the case in the BMEWS system which recently has received much publicity. These radars, although quite different from the radars which were described for the airbreathing case, do have a few things in common with the old, obsolete radars.

In general, the radars described herein have fairly discrete and separable functions. The design and development of the transmitter is accomplished to provide the proper pulse modulation and power output. The antenna structure which is responsible for formation of the beam, transmission of that beam out to the target and reception of the echoed energy must be considered. Another matter of consideration is the receiver system which analyzes the energy received from the target to determine as much as possible about the target. The last item to be considered is the data processing system which works on this data to calculate trajectories, position or characteristics of the target, or to perform intercepts, tracking and so on. In these individual fields, experts are hired who work more or less independently of one another, coordinating their efforts under a radar systems engineer to perform the engineering of a complete radar system. With the developments that have been taking place rapidly within the past few years, this procedure is fast becoming obsolete. As a matter of fact, with the advent of the phased array, managers at Rome Air Development Center were uncertain as to whether the development should be assigned to their antenna people, to their signal-processing and receiver people, or even to their data-processing people who would be responsible for working out some sort of a programmer for positioning of the beam. These functions are of equal importance in their contributions to the success of the phased array. The same is true when pulse-compression techniques are considered. The transmitter and the receiver must be designed as an integral

part so that the inverse functions can be performed properly both on transmit and receive in order to make maximum use of the information which may be available.

The important point is that some straightforward thought must be given to just what must be accomplished. The early radar problem was primarily a three-dimensional one; namely, what is the range, the azimuth and height of targets of interest? The resolution problem was primarily a Rayleigh resolution in that the essential requirement was the separation of two targets of equal amplitudes. Today the problem is much different. It is an n-dimensional problem, in that not only range is desired, but also range rate. Not only azimuth is desired, but also azimuth rate. In order to obtain velocity vectors, Doppler shifts are desired. Also sought is resolution or some means of overcoming the resolution ambiguities that occur not only in amplitude but also in time and in frequency. We must know more than just the separation of targets of equal amplitude. These may be considerably unequal in amplitude but separating them is not sufficient as, for example, in the case of a nose cone and the decoy. They might be equal in amplitude, but the question remains—which one is the nose cone, which one has the Toni? Signature effects and threat evaluation are major parts of the problem of detection and warning. Therefore, the design of the future radar will have to be accomplished by the "systems engineer." He will be required to have a clear outlook as to the over-all problem to be solved in order that he can integrate the various types of functions that must be performed and the proper types of pulse shape that suit his particular problem. Many possible shapes of pulses can be used in present radars. It very well may depend upon the specific problem to be solved and, even then, there may not be an optimum for that problem. Therefore, looking into the crystal ball, it is seen that techniques are available today, modern techniques, which threaten to completely change the radar as a system.

When one examines very carefully the future environment of radar and considers the ranges of operation, the power levels required, the resolution desired and the flexibility which is necessary, one is tempted to prophesy as to the type of radar which best meets these requirements. It is concluded that the future radar must be designed around the principles of an electronically steerable phased array. Interestingly enough, the old AN/SCR-270 was a phased array which was not electronically steerable, since the phases of the elements were fixed.

One of the very latest experimental radars to go on the air is an electronically steerable array, shown in Fig. 12. It was designed and built by the Bendix Corporation under Air Force and ARPA sponsorship to work out some of the technical difficulties in the design of such a radar.

The phased-array radar is the only new technique which offers extreme flexibility, long-range operation

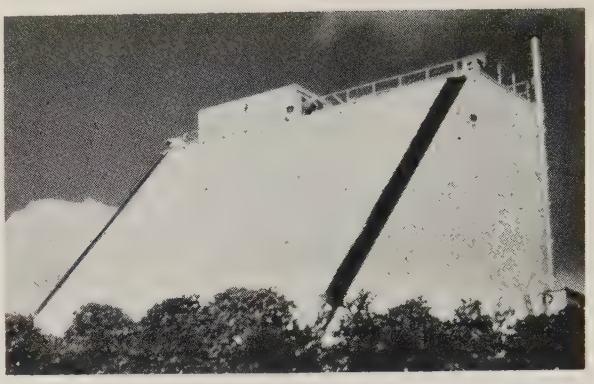


Fig. 12—Phased-array radar.

(out to ranges, beyond which means other than radar must be developed for economical reasons), very high power capability (without the usual problems associated with the generation and transmission of high power from single sources), and, very important, a high-reliability factor. The only negative factor of the phased-array radar is its expense; however, production techniques should eventually be able to reduce the cost considerably.

In conclusion, four thoughts are presented for consideration. First, the design of a radar is no longer a three-dimensional system; it is now an n-dimensional problem wherein as high as eight or more of these various dimensions must be obtained. Discrimination, resolution and signature effects are gaining in relative importance as the missile and space age is entered. Secondly, the approach to radar design must be changed to take into consideration the importance of these n-dimensions to the mission of the radar. It no longer will be possible to segregate parts of a radar for specific design and specialization, since the various aspects of the design of a radar are too closely interrelated. Thirdly, the frequency and power-level requirements are rising higher and higher; in frequency—even to the infrared, optical and ultraviolet. Several proposals recently have been presented for the development of an optical radar, particularly since the introduction of the laser. The need for increased frequency and power was presented when surveillance became a space problem, requiring very large distances as well as a capability for moving structures off the surface of the earth. Lastly, the brute-force approach in the design and development of the World War II radar—that is, to be bigger and better but more of the same—is no longer sufficient to accomplish the job which is required. It is believed that the approach of electronically steerable phased arrays is the only technique available today which promises the flexibility and capability of coping with the new space age environment.

Several technical approaches toward the solution of some of these problems have been pointed out. Many more are being studied in our various laboratories. Many more will be required if the military structure of the country is to keep ahead in this technological race.

High-Power Traveling-Wave Tubes for Radar Systems*

J. A. RUETZ†, MEMBER, IRE AND W. H. YOCOM†, MEMBER, IRE

Summary—**Data obtained on high-power traveling-wave tubes at *S* and *C* band indicate performance suitable for the final amplifier in a wide bandwidth phase coherent radar transmitter. Static phase measurements demonstrate the capability of parallel operation of tubes and the feasibility of use in pulse compression systems. Gain and power output data show that for the most efficient operation a programmed drive is needed. Data taken at constant drive power displays a compressed gain variation and a somewhat lower output. An extrapolation of present data indicates the feasibility of obtaining 25- to 30-per cent bandwidth in high-power traveling-wave tubes with other improved characteristics.**

Introduction

THE ADVENT of the high-power klystron amplifier made possible a new concept in the design of radar systems—that of phase coherence. The traveling-wave tube (TWT) advances the possibilities in radar design by providing bandwidth not previously available in high-power tubes, while maintaining many of the advantageous characteristics of the klystron amplifier. The bandwidth capability of the TWT is unsurpassed by other present day amplifying devices and hence should be of paramount interest to the modern-day radar-systems engineer.

This paper specifically discusses the characteristics of two similar high-power TWT's designed to operate in the *S* and *C* radar bands. In addition some background information is presented explaining the operation and performance related to that of the high-power klystron amplifier. It is assumed that the reader is familiar with the performance and operation of the klystron.

The characteristics of the tubes that are discussed are typical of what might be expected of the performance and operation of high-power TWT's in general. It is possible however to design for special characteristics. The influence of some of the design factors on performance is presented.

Description of the High-Power Traveling-Wave Tube

The TWT designed to operate at the megawatt power level can be compared best, perhaps, with the high-power klystron in its behavior and operation. The TWT uses similar or identical designs for the electron gun, focusing magnetic field, collector and windows. The interaction circuit however is quite different from the klystron circuit and accounts for the basic differences between the TWT and klystron.

Since the formation of the electron beam, the focusing of the beam and its collection are very similar in the TWT and the high-power klystron, it is also to be expected that the operating characteristics of the TWT and klystron would be very similar in this regard. It is true therefore that the characteristics of the modulator and associated power supplies are quite similar for the two types of tubes; any differences being a consequence of the specifications on RF performance.

The interaction circuit generally utilized for the high-power TWT is of the coupled-cavity type.[1] Coupled cavities comprise a periodic system giving rise to propagating bands (pass bands) and nonpropagating bands (stop bands). Operation of the TWT is accomplished at frequencies inside a propagating region of the coupled-cavity system. Energy is transmitted along the structure, and in the absence of the electron beam, is essentially unchanged in amplitude.

The cavities of the klystron amplifier are uncoupled except by the electron beam and, hence, the electron beam provides the phase relationships between cavities. For the TWT, however, the fields coupled between cavities by the electron beam add to the fields coupled by the cavity-coupling mechanism. The phasing of the total fields between cavities is therefore determined by the combination of the electron beam and the cavity-coupling mechanism. It follows that the TWT will be less susceptible to phase modulation than a similar klystron of equal phase length.

As would be expected from the preceding argument, gain is obtained in the TWT when the fields coupled by the cavity-coupling mechanism add approximately in phase with the fields coupled by the electron beam. It is anticipated that the TWT will be more sensitive in gain to changes in beam voltage than the klystron amplifier.

The gain mechanism is seen to involve a phasing characteristic between the wave propagating on the interaction circuit and the electron beam. In addition the magnitude of the coupling between circuit and beam is a function of the circuit fields causing motion of the electrons in the electron beam. These two factors in combination with the physical characteristics of the electron beam determine the RF characteristics of the TWT.

Proper design of the microwave interaction circuit involves obtaining the proper phasing between cavities as a function of frequency while maintaining a high

* Received by the PGMIL, January 19, 1961. The work on the VA-126 was sponsored by the Air Force under contract from the Rome Air Dev. Center, Contract No. AF 30(602)-1774.

† Varian Associates, Palo Alto, Calif.

[1] A survey article presenting some of the various forms of coupled cavity circuits is E. J. Nalos, "Present state of art in high power traveling-wave tubes," *Microwave J.*, vol. 2, pp. 31–38; December, 1959.

value for the circuit interaction fields. Several designs[2–4] have been used, each with a particular advantage.

The coupled-cavity circuit utilized for the VA-125 and VA-126 is the so-called cloverleaf circuit.[3] The circuit gets its name from the shape of the cavity. (See Fig. 1.) The coupling between cavities is obtained by means of slots in the cavity walls. The cloverleaf shape combined with the position of the coupling slots results in an increasing phase shift with frequency. This characteristic is of the type needed for a forward wave amplifier.

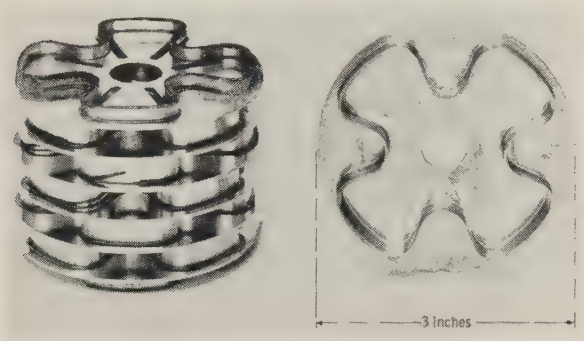


Fig. 1—The cloverleaf coupled cavity circuit used in the VA-125 and VA-126. The circuit shown is designed for the VA-126.

The propagating nature of the TWT interaction circuit makes it possible for energy reflected at the tube output to propagate back to the input. A reflection at the input will in turn mix the returned signal with the input signal and thus affect the phase and amplitude of the amplified signal. The magnitude of the phase and amplitude variations are dependent upon: 1) the magnitude and phase of the reflections present at the input and output of the device, 2) the phase length between reflections and, 3) the gain of the TWT.

In the design of TWT's it is customary to terminate the interaction circuit in one or more locations, thus lowering the necessary net gain between terminations. In this manner, it is possible to isolate input from output and reduce the effect of variations in load upon the gain and phase of the amplifier. The effect of the multiple reflections in the TWT gives rise to ripples in the small-signal gain and phase as a function of frequency. The gain ripple will be compressed if the tube is operated near saturation. However, the phase ripple will not be compressed and may even be increased. It is possible to improve the small-signal gain and phase ripple by improving the terminations. Since these variations are a function of the product of two reflections, it is possible to obtain low values of ripple by lowering the reflected power from either or both circuit terminations to near zero.

The output interaction circuit for a TWT is in general quite similar to the input section of the circuit; *i.e.*, the circuit is periodic. For high-power TWT's, it has been found advantageous to lower the phase velocity near the output to obtain increased efficiency.[5] This so-called "velocity tapering" increases the efficiency by creating better phasing conditions between the beam and circuit fields near the tube output. Resorting to nonperiodic interaction circuits increases the possibilities in TWT design and might be compared to "stagger tuning" used for broad-banded klystron amplifiers.

The foregoing qualitative description of the TWT is presented to give the reader some insight into the external factors involved in the performance of the high-power TWT. For a more complete and quantitative description, the reader is referred to the literature.[6,7]

Description of the VA-125 and VA-126

The VA-125 and VA-126 are designed to operate in the *S* and *C* radar bands. The VA-125 is made in two models—the VA-125A covers the band 2.60 to 2.95 kMc and the VA-125B covers the band 2.90 to 3.25 kMc. The VA-126 is constructed to operate in the 5.4- to 5.9-kMc band. Both tube types use the cloverleaf interaction circuit shown in Fig. 1.

A schematic cutaway view of the VA-125 is shown in Fig. 2. The electron beam is formed from a modified Pierce-type convergent gun. The electrons pass through the anode and traverse the interaction circuit to the collector. The beam is focused by a magnetic field that is uniform in the interaction region. The collector is shielded from the magnetic field and therefore the beam is spread radially before collection.

The input signal is introduced on the interaction circuit through the input coaxial window and input coupling transition. The signal is coupled from cavity to cavity, and propagates along the circuit. The interaction between the electron beam and the circuit fields causes the RF energy to experience an exponential increase with distance along the circuit.

The termination near the center of the interaction circuit absorbs the energy present on the input circuit. The signal is carried across the termination space by the electron beam and is reintroduced on the output interaction circuit. The output circuit is also terminated at

[2] M. Chodorow, E. J. Nalos, S. P. Otsuka, and R. H. Pantell, "The design and characteristics of a megawatt space-harmonic TWT," IRE Trans. on Electron Devices, vol. ED-6, pp. 48–53; January, 1959.

[3] M. Chodorow and R. A. Craig, "Some new circuits for high power TW tubes," Proc. IRE, vol. 45, pp. 1106–1118; August, 1957.

[4] M. A. Allen, "Coupling of Multiple Cavity Systems," Microwave Lab., Stanford University, Stanford, Calif., Rept. No. 584; April, 1959.

[5] J. A. Ruetz, D. Robinson, and J. Pavkovich, "The Effect of Tapered Circuits on Efficiency for High Power Traveling Wave Tubes," paper presented at PGED Meeting, Washington, D. C.; October, 1960.

[6] J. R. Pierce, "Traveling-Wave Tubes," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.

[7] W. R. Beam, D. J. Blattner, "Phase angle distortion in traveling-wave tubes," *RCA Rev.* vol. 17, pp. 86–99; March, 1956.

the center of the tube to minimize the effect of reflections at the tube output.

The signal present on the interaction circuit at the output of the tube is coupled into a reduced height waveguide. The waveguide is tapered into standard waveguide size through which the energy is transmitted into the load. The VA-126 is similar to the VA-125 except for a difference in circuit dimensions. The circuit is also liquid cooled in the VA-126 which is necessitated by the increase in power density.

The type of beam focusing used in the VA-125 and VA-126 requires higher magnetic fields than the theoretical minimum to maintain parallel flow. The increase in field required for immersed flow is compensated for by the decrease in sensitivity of the RF performance to changes in magnetic field.

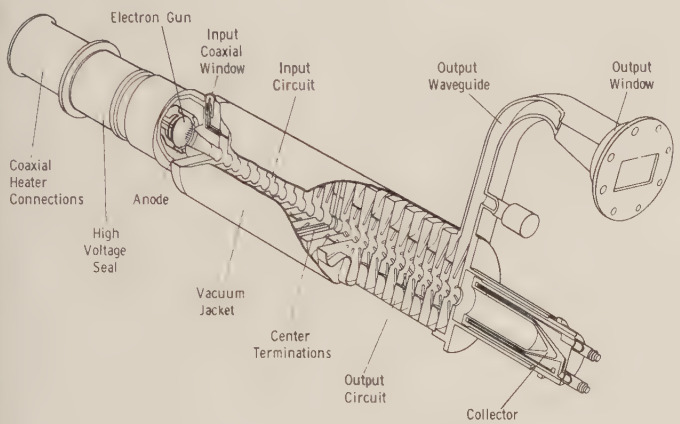


Fig. 2—A cutaway view of the VA-125 showing the internal construction of the tube.

Gain and Power Output Characteristics of the VA-125 and VA-126

Gain occurs in the TWT when the fields coupled between cavities by the cavity-coupling mechanism add approximately in phase with the fields coupled by the electron beam. A diagram showing the phase shift per cavity vs frequency for the cloverleaf circuit, as used in the VA-125A, is shown in Fig. 3 (next page). A line drawn from the origin at a constant slope on this diagram represents a constant velocity and indicates the coupling phase for the electron beam. This type of diagram can therefore be used to indicate synchronism or in-phase coupling between the beam and circuit fields.

The constant velocity line drawn in Fig. 3 is shown at a velocity corresponding to the rated beam voltage of the VA-125. A beam voltage lower than the rated beam voltage would correspond to a velocity line of less slope than that shown, while a higher voltage would have a velocity line of increased slope. It is seen that these conditions will cause the gain characteristic to be shifted toward the high-frequency end of the band for lower than rated voltage and toward lower frequencies for higher voltages. This behavior is generally true for all TWT amplifiers.

Typical saturation gain curves are shown in Fig. 4 for the VA-125A and Fig. 5 for the VA-126. The gain is seen to vary approximately 10 db over the band, and has the predicted response as a function of beam voltage.

A typical saturation gain curve for the VA-125A showing the fine structure is given in Fig. 6. The peak-to-peak amplitude ripple is seen to be approximately 3.5 db from a smooth curve. The ripple is also shown to comprise approximately $3\frac{1}{2}$ cycles across the frequency band. The data for the VA-125B and the VA-126 are similar to this. For the VA-126, $2\frac{1}{2}$ cycles of gain ripple are measured for the frequency band 5.4 to 5.9 kMc. This value for the number of cycles of ripple agree with calculation assuming multiple reflections between the input of the interaction circuit and the circuit termination, and also between the circuit output and the circuit termination. Mismatches of the load would increase the number of cycles and also change the ripple amplitude.

No attempt has been made to flatten the gain variation with frequency. This can be accomplished in several ways and will be discussed in a later section.

Typical saturated output power data for the VA-125A and VA-126 are presented in Figs. 7 and 8. This data is taken with drive power adjusted for maximum output. The power output is seen to peak toward the lower frequencies as the beam voltage is increased. The maximum efficiency is nearly unaffected by beam voltage.

The gain and peak power output data indicates acceptable performance for a radar operating at a fixed frequency or for a radar for which drive power can be adjusted. For constant drive operation a decrease in power output must be accepted at some frequencies. Data taken on the VA-125B for constant drive power is shown in Fig. 9. The reason for the measured decrease in peak power output under these conditions is that the frequency region giving the highest peak power output also has the highest gain. To obtain maximum frequency coverage, the tube must be driven past saturation in the band center which decreases the power output. The tube is generally underdriven at the edges of the band and hence less power is also obtained at these frequencies.

The efficiency calculated from the data presented in Figs. 7 and 8 results in a maximum value of 40 per cent obtained on the VA-125 and greater than 35 per cent on the VA-126. These values of efficiency exceed those obtained from broad-band klystrons of comparable bandwidth and power level. The TWT has not attained the efficiency of the narrower bandwidth klystrons however.

Average powers of 7.5 kw have been obtained for both the VA-125 and VA-126. The average power levels have been limited by the power handling capability of the internal termination. The eventual limitation on average power level is expected to be similar to the klystron.

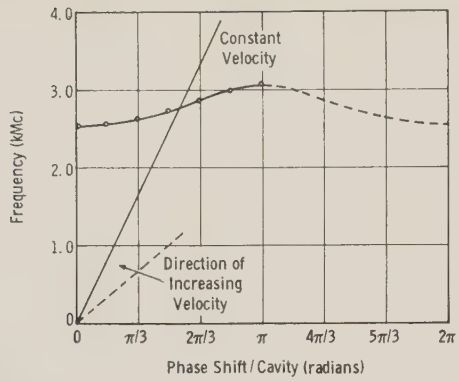


Fig. 3—Phase shift per cavity vs frequency for the cloverleaf circuit used in the VA-125A.

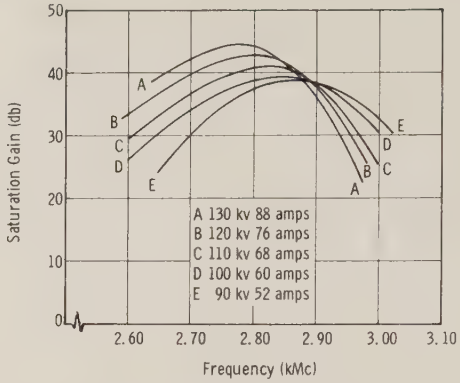


Fig. 4—Typical saturation gain curves for the VA-125A. The ripples in the gain curves are not shown.

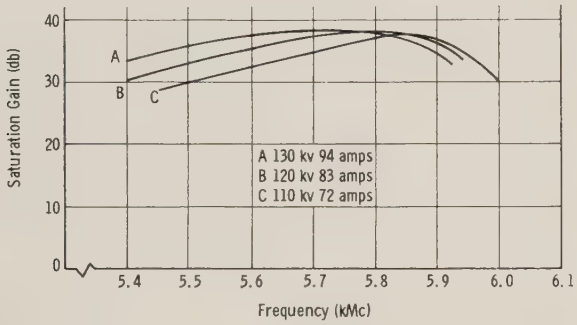


Fig. 5—Typical saturation gain for the VA-126. Gain ripples are not shown.

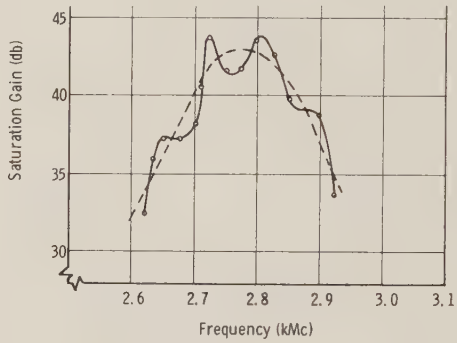


Fig. 6—Saturated gain vs frequency showing the typical fine grain structure for the VA-125A.

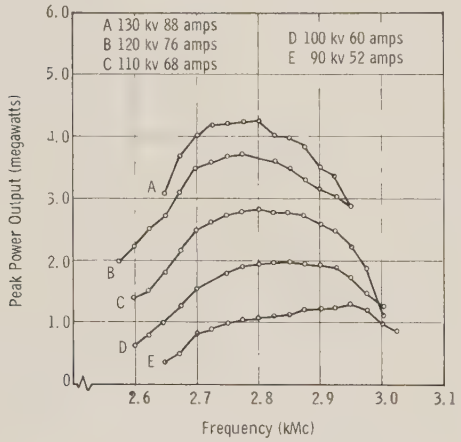


Fig. 7—Typical saturated power output for the VA-125A.

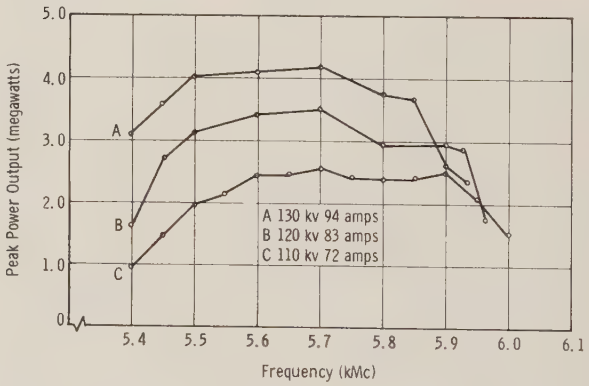


Fig. 8—Typical saturated power output for the VA-126.

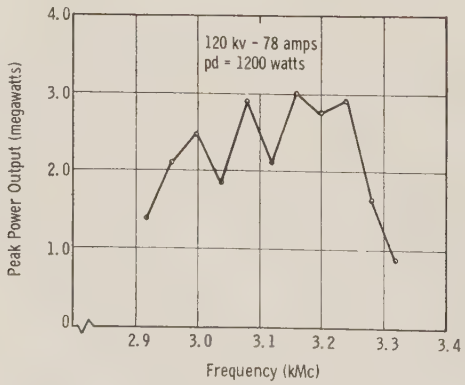


Fig. 9—Typical peak power output for the VA-125B. The data is taken under constant drive power conditions.

Phase Characteristics of High-Power TWT's

The phase characteristics of the final amplifier are of paramount importance in the design of modern radar systems. Phase deviation from linearity, phase sensitivity to beam parameters, and phase variation with drive power are the factors of most interest.

Deviations from linear phase were suggested previously as occurring because of multiple reflections between sections of the interaction circuit. Since the interaction circuit is coupled to the load, mismatches in the load can also cause reflections which in conjunction with the internal termination of the circuit will affect the gain and phase response of the TWT. Again it should be noted that a perfect internal termination will eliminate effects of the load on the gain and phase response of the amplifier.

A second effect on phase linearity is caused by variations in input drive with frequency or similarily a variation in gain of the amplifier for constant drive power. As energy is extracted from the beam, the net velocity of the electrons is reduced and hence phase changes can occur. The magnitude of this phase change will be dependent upon the strength of the interaction fields and upon the parameters of the beam.

Data on phase deviation as a function of input power level taken from static measurements on the VA-125B are shown in Fig. 10. It is seen that a maximum phase deviation of 30° is measured as a result of a 20-db change in input power.

Phase deviation from linearity data as a function of frequency taken on the VA-125B under constant drive power conditions are plotted in Fig. 11. A sample of input power and output power were compared by means of a phase bridge in making these measurements. The power input was adjusted to be constant over the frequency band. A repeat of the phase linearity measurements gave results essentially the same as those shown. Due to the many variables involved in this measurement however it was not possible to reproduce the data to any great accuracy.

Approximately the same number of ripples are present on the phase linearity characteristic as was obtained from the measurement for gain. The gain variation across the tube bandwidth is probably the reason for the negative phase deviation at the edges of the band.

Measurements on phase sensitivity to beam voltage are given in Fig. 12, also taken on the VA-125B. These data are shown as a function of beam voltage with frequency as a parameter. The variation of an equivalent phase length klystron is approximately twice the measured value for the TWT.

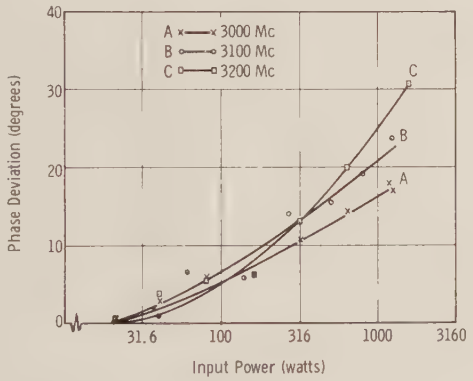


Fig. 10—Phase deviation as a function of drive power for the VA-125B.

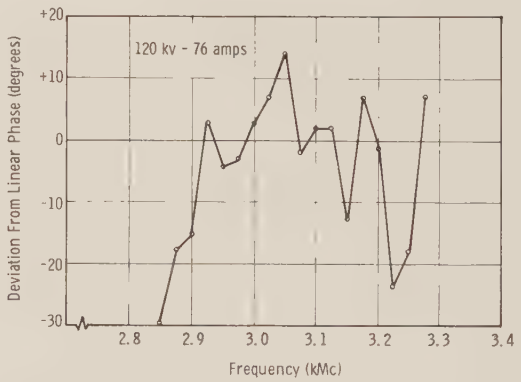


Fig. 11—Deviation from phase linearity for the VA-125B. Input power is constant at 1 kw.

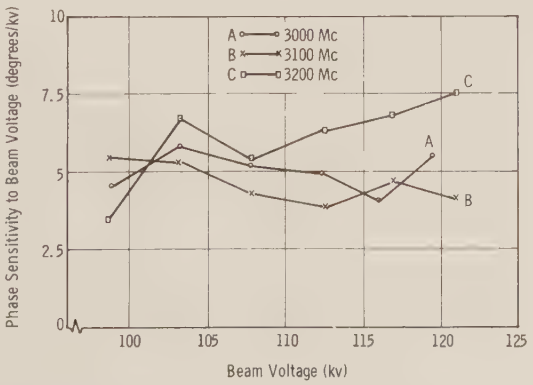


Fig. 12—Phase sensitivity vs beam voltage for the VA-125B. Input power is constant for maximum power output at 120-kv beam voltage.

Nonlinear Effects in High-Power TWT's

Like other power amplifiers the TWT operates at maximum efficiency in a nonlinear region. The nonlinearity of the TWT is similar to that of the klystron in that the highly bunched beam contains harmonics of the fundamental frequency. This harmonic beam power does not in general interact strongly with the circuit fields in the high-power TWT. However some energy is coupled into the output waveguide. A similar effect exists also for the klystron amplifier, energy being coupled into the output cavity at harmonic frequencies.

A second nonlinear effect of interest is the mixing that occurs when the TWT amplifier is driven by two or more input signals. The signals of interest in this type of operation are those near the frequency band being amplified. These signals are generated as a result of mixing between the fundamental and harmonics of one signal with harmonics of another signal. Likewise the fundamental and harmonics of the second signal mix with the harmonics of the first signal.

A third nonlinear effect is the compression that occurs to one signal as a second signal is increased in amplitude. In general, as compression occurs, mixing builds up and other signals become noticed in the amplifier output.

The measurement of nonlinearities including those of harmonic output is not easily made. To make harmonic measurements at the output of the TWT, it is necessary to measure each mode of propagation in the output waveguide. It is possible to simplify these measurements somewhat by investigating the output system of the amplifier. It was stated earlier that the output guide for the VA-125 and VA-126 was of reduced height. The height used in these tubes is such that only two modes can propagate at the second harmonic frequency. In addition the symmetry of the coupling mechanism assures that only one of these modes is excited. If the waveguide system is uniform and transfer of energy between modes does not occur, then this one mode only need be measured. The mode of course that remains is the TE_{10} mode.

The results of the second harmonic measurements are presented in Table I and are seen to verify the above statement. The measurement of power in the TM_{11} mode at 5.40 kMc appears to be the major exception to the statement that the second harmonic power propagates only in the TE_{10} mode. However this can be explained by the possibility of a small error in the calibration of the probes which is enough to misidentify the TM_{11} mode.

Measurement of third harmonic power for any of these tubes has not been made. This measurement is more complicated than the second harmonic measurement and is generally of less interest. Measurements of second harmonic power present in the output of the VA-125B and VA-126 have also not been made. However it is expected that the results would be similar to those obtained on the VA-125A.

The variation of second harmonic power with frequency is possibly explained by coupling occurring between the second harmonic currents in the electron beam with higher frequency modes in the cloverleaf circuit. A mode exists at the second harmonic frequency that does synchronize with the beam near the frequency region having the largest amount of second harmonic power. More measurements are needed for verification of this effect.

Measurement of cross-modulation effects is also difficult to perform. A correct method for this measurement involves using an input signal having a broad noise spectrum with a channel or slot missing in the band. The measurement then involves measuring the noise amplitude in the slot as a function of noise amplitude over the remainder of the band.

No measurements have been made of cross-modulation effects in high power TWT's. A measurement had been planned which involved driving the high-power TWT at two frequencies simultaneously. A measurement of the output spectrum would then give an indication of cross-modulation effects. The difficulties involved in measuring relative amplitudes of pulsed spectrums combined with the cross-modulation power existing at the output of the driver chain made this measurement unfeasible.

TABLE I

Second Harmonic Power for the VA-125A*

Mode	Frequency 5400 Mc	Frequency 5600 Mc	Frequency 5700 Mc
	Power Below Fundamental (db)	Power Below Fundamental (db)	Power Below Fundamental (db)
TE_{10}	29.2	26.8	39.2
TE_{20}	36.1	50.0	40.5
TE_{01}	58.0	$\propto$	64.6
TE_{11}	36.4	49.8	61.8
TM_{11}	21.7	44.5	57.0
Total	20.7	26.6	36.7

* The drive power was adjusted for maximum power output at the fundamental frequency.

Future Developments in High-Power TWT's

The results obtained on high-power TWT's are seen to fulfill some of the characteristics needed for radar applications. Several of these characteristics can be improved by proper design. Some, however, are characteristic of the traveling-wave tube interaction process and not subject to control. Characteristics not controllable are the phase sensitivity of the amplified signal to beam-voltage changes and the variation of phase shift with input drive level. Characteristics such as gain and phase ripples measured on the VA-125 can be minimized since they are caused by imperfect terminations of the interaction circuit.

The present highly efficient performance has been

obtained for the VA-125 and VA-126 by utilizing a tapered velocity of the interaction circuit near the tube output. The shape of the gain curve can also be controlled by proper selection of the circuit velocity along the total length of the interaction circuit. Since the circuit is terminated at the center, improvement in gain response can also be obtained by peaking the gain of the input section to a frequency removed from the peaked response of the output circuit. The utilization of these and similar measures would obtain maximum bandwidth of tubes utilizing any given interaction circuit. Increased bandwidth would be obtained by using an interaction circuit with greater cold bandwidth. Designs exist for circuits having a greater bandwidth than the cloverleaf circuit. The utilization of these will be made when the system requirements justify their use. Bandwidths of 25 to 30 per cent should be feasible.

It is expected that efforts will be made to increase the efficiency of high-power TWT's beyond that measured to date. Use of a depressed collector should help toward this end. The use of a depressed collector has disadvantages, however, in increasing the complexity of the modulator and of the tube itself. In addition the phase characteristics can be greatly altered if electrons are allowed to return to the interaction space. This latter effect is a function of proper design and could be controlled to any extent needed for a system design. Efficiencies greater than 60 per cent should be realizable with little degradation in RF performance.

Other factors, such as harmonic output and cross-modulation effects, would need further measurement and study to ascertain methods of control. The design of the interaction circuit can be made to lower the coupling of the beam at harmonic frequencies which should lower harmonic- and cross-modulation output.

The average power capability of the VA-125 and VA-126 is not expected to be a limiting factor in this type of high-power TWT. It is possible to replace the internal termination used in these tubes with external ones which are connected to the interaction circuit through matching waveguides. The external terminations can then be designed to handle the total amount of power transmittable through the waveguide system. Other factors will of course eventually limit the average power obtainable at any one frequency band. These limitations are the same as those which eventually will limit the average output power of other similar beam devices.

Acknowledgment

The measurements of second harmonic power were made by J. Rooney of the General Electric Microwave Laboratory under the sponsorship of the Rome Air Development Center.

Automatic Frequency Control of Magnetrons*

AUSTIN R. SISSON†, SENIOR MEMBER, IRE

Summary—A practical AFC system for both *X*-band and *Ku*-band magnetron radar transmitters has been developed. This system meets the needs of three radar systems currently in production. Frequency stabilization is accomplished by electronically controlling the magnetron load impedance. Therefore, the magnetron "pulling figure" is put to use.

The closed loop AFC system maintains frequency stability to better than ±2 Mc at *Ku*-band and ±1.5 Mc at *X*-band frequencies over wide environmental conditions. The system dynamically compensates for deviations in frequency due to temperature, vibration, shock, changes in duty cycle, drift with life, and poor voltage regulation. The response time of the AFC system is such that, starting at the limits of the capture range, "on frequency" response is obtained within the first five pulses of the magnetron. One of the radar systems requires a considerable delay between groups of pulses. For this radar, the AFC system provides a memory to hold the conditions for "on frequency" response during the delay periods.

A qualitative analysis of the closed-loop system is made with reference to the Rieke Diagram. It is shown how a single magnetron has been used to meet two different sets of system performance requirements.

System operation and hardware are discussed with reference to practical difficulties.

* Received by the PGMIL, January 17, 1961.
† The Bendix Corp., Bendix-Pacific Div., North Hollywood, Calif.

Introduction

NUMEROUS applications exist for a small-size lightweight transmitter with its output frequency stabilized to close tolerances. Previous work has been done using such techniques as mechanical servo, stalo cavities, and the control of system parameters which affect the frequency, such as temperature and tube input voltages. In general, these techniques have shortcomings in the areas of size and weight, and/or accuracies, and response time. The rapid growth of the art in microwave ferrite devices has opened up new

avenues for automatic frequency control in the microwave area. A practical AFC system using a ferrite device has been developed for use with both *X*- and *Ku*-band magnetrons. This AFC system is used in three airborne radar systems now in relatively large quantity production.

The block diagram of Fig. 1 shows the major components of the system. These elements include a magnetron with a large pulling figure, an electronically controlled ferrite phase shifter, a slide screw tuner, a microwave frequency discriminator, and a servo amplifier. Basically, AFC action is accomplished by operating the magnetron into a fixed VSWR with a controlled phase.

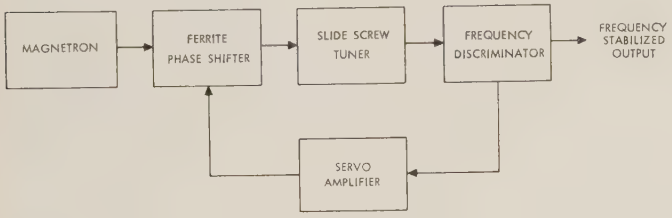


Fig. 1—Block diagram.

Initially, the system is adjusted to the desired operating frequency f_0. The sequence of operation is then as follows. The magnetron changes frequency by Δf due to either, or a combination of, the parameters which affect frequency. A portion of the transmitted energy at $f+\Delta f$ is coupled to the discriminator. The discriminator output signal proportional to Δf is delivered to the servo amplifier. This signal is amplified and converted to a driving current that controls the ferrite phase shifter. The phase of the fixed VSWR introduced by the slide screw tuner is changed by the proper amount and sense to "pull" the magnetron back on frequency. A stabilized frequency output results when the magnetron frequency is equal to the discriminator reference frequency. Therefore, the magnetron's pulling characteristic is put to use to maintain dynamically a constant output frequency.

This paper presents a qualitative discussion of the AFC system in terms of a specific radar requirement, as well as general design considerations.

Basic Considerations

An understanding of the operation of the AFC loop and an insight into the design of such a system will be facilitated by a discussion of certain magnetron and AFC system parameters with reference to the Rieke Diagram.[1] A typical Rieke Diagram is shown in Fig. 2. Essentially, this diagram presents a plot of constant power and frequency contours on a Smith chart as a function of the magnetron load impedance.

[1] G. B. Collins, "Microwave Magnetrons," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 178–187; 1948.

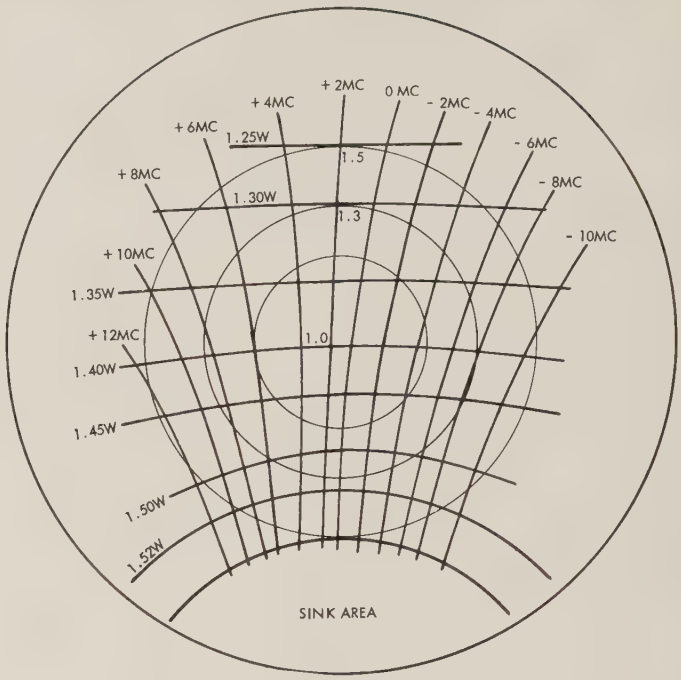


Fig. 2—Typical Rieke Diagram.

A. Pulling Figure

The pulling figure of the magnetron is usually defined to be the total frequency excursion that occurs as a VSWR magnitude of 1.5:1 is varied through 180 electrical degrees. This is equivalent to traversing 360° or completely around the Rieke Diagram.

B. Capture Range

The capture range for the AFC system of this paper is defined as the maximum frequency excursion vs phase for a given VSWR. This value, of course, is the same as the pulling figure defined for this same given VSWR. The end points are the frequency limits within which the system can compensate. The capture range can be increased by increasing the VSWR within certain limitations.

C. Phase Function

Throughout this paper, a clockwise rotation around the Rieke Diagram is defined as an increase in phase. In Fig. 2, it is seen that starting at the left of the figure and traversing a constant VSWR circle with increasing phase causes a decrease in frèquency. However, a continued increase in phase around the bottom of the diagram causes the magnetron frequency to increase. Therefore, the magnetron frequency is a double-valued function of the phase of the load impedance.

D. Sink

The "sink" area, located at the convergence of the constant-frequency curves, represents the region of instability. This region is the locus of impedance points

that produce a poor frequency spectrum, frequency jumps, moding, or non-oscillation. To produce such an impedance, a high VSWR magnitude and a definite phase are required. This area is commonly referred to as the "worst phase condition."

E. Two Modes of Operation

Fig. 3 illustrates the possibility of two modes of operation. The Rieke Diagram can be divided into two regions. Region I is characterized by decreasing frequency with increasing phase, low power output, and the possibility of increasing the VSWR magnitude to high values with stable operation. As previously mentioned, the high VSWR produces a greater capture range. Region II is characterized by the "sink" area, higher power output, increasing frequency with increasing phase, and a smaller capture range possibility than in Region I. The presence of the "sink" area in Region II necessitates maintaining the VSWR below a certain maximum value. Therefore, the maximum capture range possible in Region II is smaller than in Region I.

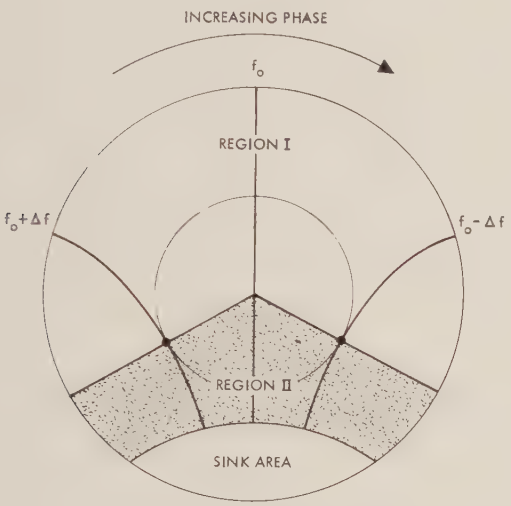


Fig. 3—Modes of stable operation.

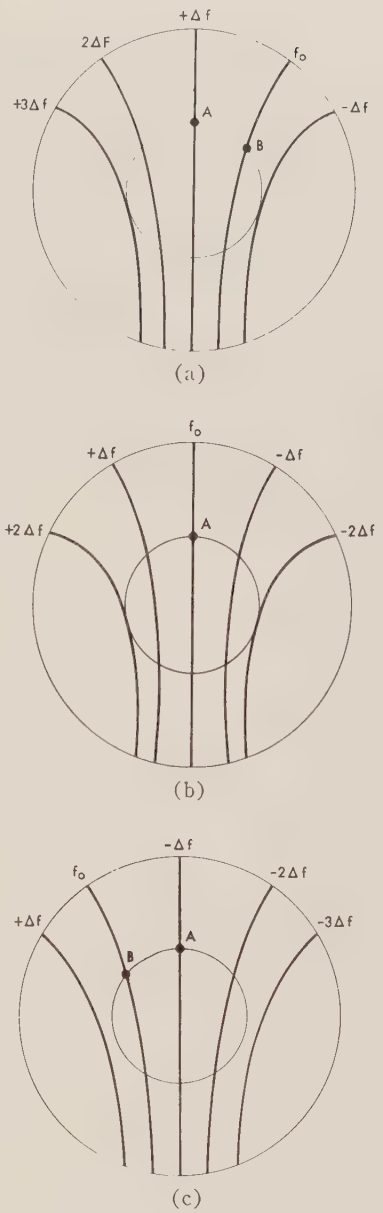


Fig. 4—Magnetron temperature characteristics: (a) cold temperature, (b) room temperature, (c) hot temperature.

F. Long-Line Effects

The frequency change is continuous with smooth change in phase. Erratic frequency changes due to long-line effects are eliminated by locating the variable magnetron load impedance at the output terminals of the magnetron.

G. Rieke Diagram vs Frequency Change

The effect of a frequency change on the Rieke Diagram can be understood by referring to Fig. 4. If the magnetron load impedance has no reactive component, the output frequency will correspond to that constant-frequency contour through the middle of the Rieke Diagram. Assume that the magnetron is initially set up at room temperature with load impedance corresponding to Point A of Fig. 4(b). The output frequency will be f_0. Now assume that the magnetron has a negative temperature coefficient, so that at a higher temperature and with the impedance remaining at Point A in Fig. 4(c), the frequency will decrease to $f_0-\Delta f$. Similarly, the frequency will change to $f_0+\Delta f$ as in Fig. 4(a) for some low temperature. However, if the phase of the VSWR is adjusted to Point B in either the high or low temperature case, the output frequency will again become f_0. On correlating this discussion with the definition of capture range, it is seen that f_0 will be outside the capture range if the temperature is increased somewhat above that shown in Fig. 4(c). In this event, it will no longer be possible to "pull" the output frequency back to f_0 by a simple adjustment of phase.

With the magnetron operating into a constant load, a change in frequency due to any effects will cause a rotation of the constant frequency contours on the Rieke Diagram. However, if f_0 remains within the capture range as defined in this paper, it will be possible by adjustment of phase only to "pull" the output frequency to f_0.

System Discussion

In this AFC system, the magnetron is operated into a constant VSWR load impedance. A ferrite phase shifter is used to vary the phase of the reflection in a controlled manner, such that the output frequency is maintained constant within close tolerances over wide environmental conditions. The same *Ku*-band tube type is operated in the two areas of the Rieke Diagram to meet the requirements of two different radar systems.

A. Component Characteristics

1) *Magnetron:* The proper choice of a magnetron, of course, must be based on a consideration of the over-all requirements of the transmitter. However, the only magnetron parameter peculiar to the usual magnetron specification is the pulling figure. Usually, in the conventional system, it is desirable to minimize this parameter. However, in this system a large pulling figure is required. The requirement for the minimum pulling figure is determined by analysis of the factors which affect frequency, taking into account the environmental conditions. The usual factors are temperature, vibration, shock, changes in duty cycle, drift with life, and poor voltage regulation. Also, in many cases, the frequency is affected by other parameters such as AC filament magnetic field. A maximum pulling figure is established compatible with tube manufacturing tolerances and the maximum AFC loop gain which can be tolerated.

Consideration should be given to the two modes of operation described above in selecting a tube with capture range and power output compatible with system requirements. On-the-shelf tubes, generally, will not have a sufficiently high pulling figure. However, existing tubes can be modified to increase the pulling figure. If a new tube development is required, consideration should be given to operation in Region I of Fig. 3. It is possible to design the tube for high efficiency in this region by moving the "sink" area up to cover essentially the entire area of Region II. This tube design also results in a large capture range.

In the present application, a new *Ku*-band tube was developed, and an existing tube was modified for use at *X* band. The *Ku*-band tube is used in two radar systems requiring different power output and capture ranges. The tube is operated in Region I for the radar requiring a high capture range and a relatively low power output. The second radar requires a higher power output, but somewhat lower capture range. The tube is compatible to the second radar when operated in Region II.

2) *Ferrite Phase Shifter and Slide Screw Tuner:* The variable magnetron load impedance consists of a slide screw tuner and an electronically controlled ferrite phase shifter. The capacitive screw provides for initial adjustment of the VSWR magnitude. The sliding feature permits mechanical phase adjustment to meet the initial conditions for operation in Region I or Region II of Fig. 3. This adjustment reduces the total phase shift requirement of the ferrite phase shifter and compensates for production tolerances in the waveguide and tube.

In general, many different phase shifter types are applicable to this AFC system. The choice of a DOFL (Diamond Ordnance Fuze Laboratory)[2] type phase shifter for this particular design was based primarily on the premium requirement of reduction in size and weight. The DOFL geometry offers the largest reciprocal phase shift per unit length. This design incorporates temperature compensation and maintains a low input VSWR over its phase shift range. Temperature compensation is accomplished by a combination of magnetic shunting and open-loop thermistor current control. A low phase shifter VSWR is required over the temperature and control current ranges in order to maintain proper control of the magnetron pulling characteristics. During the course of design of the DOFL phase shifter, considerable difficulty was encountered in obtaining a sufficiently low VSWR. However, the completed design resulted in a VSWR of less than 1.1 over the operating conditions for both *X* and *Ku* bands.[3]

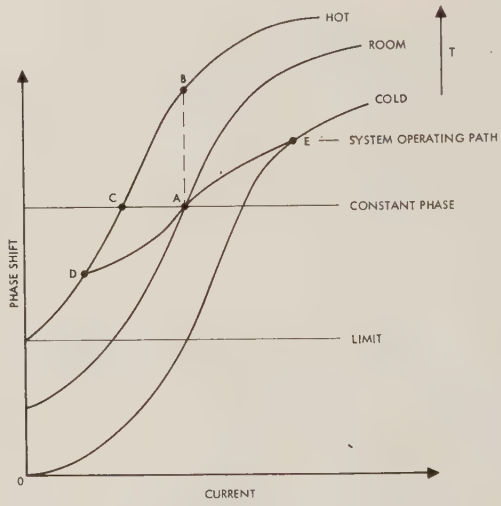


Fig. 5—Ferrite phase shifter characteristics. DAE is a typical operating phase shift vs current curve for operation in Region I of the Rieke Diagram, assuming no transients.

[2] F. Reggia and E. F. Spencer, "A new technique in ferrite phase shifting for beam scanning of microwave antennas," Proc. IRE, vol. 45, pp. 1510–1517; November, 1957.

[3] A. Clavin, "Problems Associated with Rectangular Waveguide Phase Shifters," presented at PGMTT Natl. Symp., San Diego, Calif., May 9–11, 1960. Reprints available from Rantec Corporation, Calabasas, Calif.

Fig. 5 shows a typical family of phase shift vs current curves with temperature as the parameter. With the use of thermistor current control, the increase in phase from Point A to B as the temperature is increased from room to hot, is cancelled by the decrease in phase as the current is reduced from Point A to C. The effect of proper magnetic shunting is to reduce the spread in the family of curves. In the limit, perfect magnetic shunting would result in only one curve for all temperatures.

3) *Discriminator:* The discriminator, of course, is a very important component of the AFC system, since it is the frequency reference for the system. Many discriminator techniques, also, may be used in this AFC system. However, as in the case of the phase shifter, the choice of the discriminator for this particular design was weighted heavily by the size and weight requirements.

The discriminator technique selected consists of a cross-guide directional coupler used to couple a small portion of the power from the main line. This power is fed to a magic tee used as a power divider with isolation between its two outputs. An assembly consisting of a miniature band-reject type cavity followed by a crystal detector mount is connected to each of the two outputs of the magic tee. The cavities are temperature compensated and mounted in a constant pressure area. One cavity is tuned slightly above and the other slightly below f_0. The crystal outputs are fed to a differential amplifier, the output of which is the familiar "S" curve. The discriminator output is in the form of positive or negative pulses, depending upon the sense of the frequency change. Cross-over accuracies of ± 0.5 Mc at Ku band and ± 0.25 Mc at X band have been obtained.

The cavities must be temperature compensated and either sealed or mounted in a dry, constant-pressure area in order to meet the cross-over accuracies quoted. A small change in air pressure will cause a considerable drift in cross-over frequency.

Considerable difficulty was encountered during the design phase of the Ku-band system due to poor crystal stability over the temperature range. Crystal selection by means of heat cycling tests resulted in a low yield of crystals which would meet the cross-over accuracy requirements. The availability of an improved crystal relieved this area, although some crystal selection is still required. No great difficulty in this area was encountered with the X-band system.

4) *Servo Amplifier:* The servo amplifier consists of an integrator circuit and a dc power amplifier. These circuits convert the pulse input from the discriminator to a controlled dc output. The integrator converts the pulses to dc, and provides memory to maintain "on frequency" response conditions during short standby periods of the radar transmitter. A current-limiting circuit in the output of the power amplifier is used to limit the phase-shifter control current on the upper end of the range. This circuit is the electrical stop referred to in another section of this paper. A gain adjustment is provided for initial alignment of the system.

B. Operation in Two Regions of Rieke Diagram

In order to obtain proper operation in Region I, the AFC loop must be interconnected in order to provide a decrease in frequency with an increase in phase. Adjustments then begin with an open loop and with zero control current to the phase shifter. By means of the slide screw tuner, the impedance is adjusted to Point A of Fig. 6. This point is the upper limit of the capture range. Now, as the control current is increased, the impedance moves clockwise along the constant VSWR circle through Region I to Point B with a consequent decrease in output frequency. An adjustable electrical stop is provided to limit the control current to less than that value which would increase the phase beyond Point B and into Region II. This stop is adjusted and the AFC loop is closed. Upon closing the loop, the error signal is fed around the loop to the phase shifter in the proper sense in order to adjust the phase to Point M and the magnetron is on frequency (f_0). Now, should environmental changes or other factors tend to cause a frequency change, a rotation of the constant frequency contours occurs as shown in Fig. 7. Under these conditions, the AFC system automatically and continuously controls the load impedance phase, such that the impedance is maintained at Point P with a resultant constant output frequency. Point M of Fig. 7 is the impedance which was established initially. Under constant impedance conditions, the output frequency would have been $f_0 - \Delta f$.

Operation in Region II may be described similarly. However, the loop must be interconnected so that an increase in phase causes an increase in frequency. Also, the VSWR must be reduced in order to stay out of the "sink" area. Operation along the VSWR circle between the Points C and D of Fig. 6 is typical in Region II. The Points N and O of Fig. 7 are the initial and "on frequency" impedances, respectively.

When the radar system is in the standby condition, the magnetron is not operating. Therefore, there is no error signal from the discriminator. Under this condition, the servo amplifier will drift and drive the phase shifter to one of its limits. When the radar is activated, the action is similar to that described on closing the loop after initial adjustment.

C. System Performance

The frequency stabilization accuracies of the specified design in this paper are ± 2.0 Mc at Ku band and ± 1.5 Mc at X band. The AFC system meets the radar requirement of being on frequency within the first five pulses of the magnetron. The corresponding time is approximately 30 milliseconds. This response was measured with initial conditions such that the frequency of the first pulse was at the limit of the capture range.

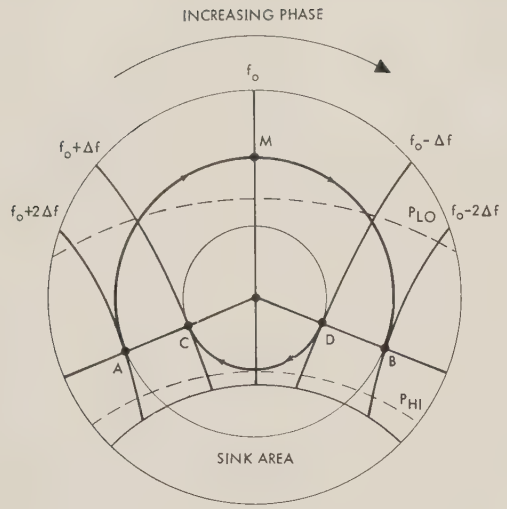


Fig. 6—System operations in Regions I and II.

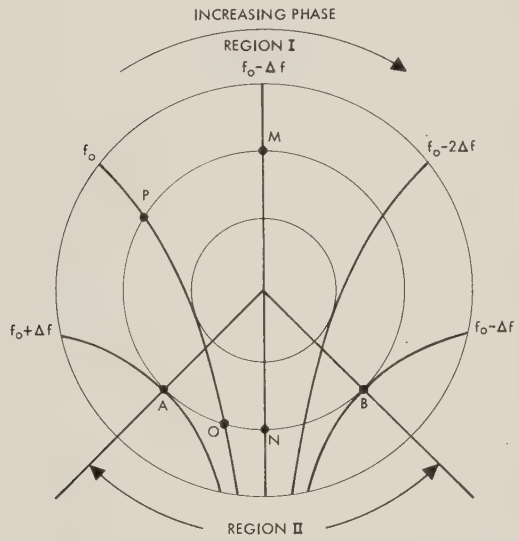


Fig. 7—System operation under environmental conditions.

Conclusion

A method for maintaining an automatic frequency control of magnetron output has been presented. The output impedance of the magnetron is electronically controlled to stabilize the frequency. This practical method is being used in three radar systems currently in production. The stabilized output frequency has been held to ± 1.5 Mc at *X* band and ± 2.0 Mc at *Ku* band over a wide environmental range.

Several terms have been described and applied to the basic theory underlying this AFC technique. Illustrations have been given of the pulling figure, sink area, capture range, phase ambiguity, and the frequency vs load impedance concept.

System operation and a qualitative analysis have been presented. This analysis should enable the interested engineer to observe these results using the basic components of the system.

It has been shown that two regions for stabilized frequency operation exist. A single magnetron has been utilized in each of these regions to meet the specific requirements of two individual systems. In Region I, a large capture range is available to meet wide environmental variations at a reduced power. In Region II, a power output is available over a lower environmental range.

It is the author's opinion that the AFC technique may be used with variations to solve several problem areas requiring stable high-power oscillations.

Acknowledgment

The author wishes to express appreciation for the assistance of his associates. D. B. Weller was extremely helpful in discussions of the work and preparation of the text.

The Radar Measurement of Range, Velocity and Acceleration*

E. J. KELLY†

Summary—This paper is a study of the ultimate attainable accuracy in the radar measurement of range, range rate, and range acceleration. It is assumed that these quantities are to be measured by a coherent radar with a large output signal-to-noise ratio. The approach is entirely theoretical, and the accuracy evaluated is the accuracy that would be attained with an ideal receiver which performs maximum-likelihood estimates of the unknown parameters. The transmitted waveform is fixed and arbitrary, and the error variances and covariances are evaluated in detail in terms of the amplitude and frequency modulation of the transmitted wave. Specific results are also given for constant amplitude pulses carrying arbitrary combinations of linear and quadratic frequency modulation.

Introduction

THE general theory of radar measurement[1,2] is briefly reviewed and is then applied to the measurement of acceleration together with range and velocity. Of course, it is actually range, Doppler frequency and Doppler rate which are measurable in a monostatic radar, and the familiar ambiguity function for range and Doppler measurement is extended to an ambiguity function for all three quantities. The moment matrix for the measurement errors in these three quantities is computed for the strong-signal (small-error) case, and the effects of various types of amplitude and frequency modulation are discussed. An interesting feature is the dependence of the velocity error on the true acceleration and an unavoidable, acceleration-dependent correlation between the range and velocity errors. This paper extends the results of Manasse[3] and Bello,[4] whose analyses apply specifically to pulsed radars.

Detection and Parameter Estimation

Suppose, at first, that a definite narrow-band signal, $\sigma(t) = \text{Re}\left\{s(t)e^{i\omega_0 t}\right\}$, is to be detected in a background of white noise. Here ω_0 is the angular carrier frequency and $s(t)$ is the complex signal modulation. The background noise can be represented in the analogous form $\xi(t) = \text{Re}\left\{n(t)e^{i\omega_0 t}\right\}$, where $n(t)$ is a complex, white Gaussian process. We have, formally,

$$E\overline{n(s)}n(t) = 2N_0\delta(t-s), \tag{1}$$

where N_0 is the noise power per cycle (one-sided) of the noise $\xi(t)$. (E stands for expectation, or ensemble average, and the bar denotes the complex conjugate.) If it is supposed that $\sigma(t)$ appears in the noise with unknown amplitude A and unknown carrier phase δ, then the detection problem consists of deciding whether the complex modulation $z(t)$, which is actually present on the carrier is simply noise $n(t)$, or signal-plus-noise, $Ae^{i\delta}s(t)+n(t)$.

The well-known matched-filter solution to this problem is to observe the envelope of the output of a matched filter at a specified time. If this quantity is sufficiently large, the signal is detected. Analytically, the detection criterion is

$$\left|\int_{-\infty}^{t_0}\overline{s(t)}z(t)dt\right| > \text{const.} \tag{2}$$

If the signal is of finite duration, and time is available to process it, the upper limit t_0 can be replaced by $+\infty$. Note that the mean-square noise output of the matched filter is

$$\begin{aligned} E\left|\int_{-\infty}^{\infty}\overline{s(t)}n(t)dt\right|^2 &= E\int\int_{-\infty}^{\infty}\overline{s(t)}s(t')n(t)\overline{n(t')}dtdt' \\ &= 2N_0\int\int_{-\infty}^{\infty}\overline{s(t)}s(t')\delta(t-t')dtdt' \\ &= 2N_0\int_{-\infty}^{\infty}|s(t)|^2dt. \end{aligned}$$

The (peak) signal output is just

$$A\int_{-\infty}^{\infty}|s(t)|^2dt,$$

so that the output signal to rms noise voltage ratio is

$$\frac{A}{\sqrt{2N_0}}\left\{\int_{-\infty}^{\infty}|s(t)|^2dt\right\}^{1/2} = \sqrt{\frac{E_s}{N_0}}, \tag{3}$$

where E_s is the total energy contained in the signal.

Now suppose that the expected signal is not known precisely, but is only known to be one of a family of signals, whose complex modulations are a given function, $s(t;\alpha_1\cdots\alpha_n)$, of n real parameters $(\alpha_1\cdots\alpha_n)$. We are to detect the presence of such a signal and simultaneously to measure (estimate) the values of the signal parameters $(\alpha_1\cdots\alpha_n)$. The logical procedure is to provide a set of matched filters, one for every set of parameter values, and to observe the output SNRs.

* Received by the PGMIL, January 19, 1961. Operated with the support of the U. S. Army, Navy, and Air Force.

† Lincoln Lab., Mass. Inst. Tech., Lexington, Mass.

[1] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise. I," *J. Soc. Ind. Appl. Math.*, vol. 8, pp. 309–341; June, 1960.

[2] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise. II," *J. Soc. Ind. Appl. Math.*, vol. 8, pp. 481–507; September, 1960.

[3] R. Manasse, "Parameter Estimation Theory and Some Applications of the Theory to Radar Measurements," Mitre Corp., Bedford, Mass., MITRE Tech. Ser., Rept. No. 3; April, 1961.

[4] P. Bello, "Joint estimation of delay, Doppler, and Doppler rate," IRE Trans. on Information Theory, vol. IT-6, pp. 330–341; June, 1960.

The parameter values associated with the filter with maximum output SNR provide the desired measurement, while the value of this output can be used to decide on the actual presence of a signal.

It can be shown that this combination of detection and measurement is equivalent to the general statistical method of maximum likelihood decision. On the other hand, the simple, intuitive, matched-filter approach can readily be generalized to nonstationary, non-Gaussian noise. In practice, one usually finds a way of simplifying the processing; a finite number of filters is used (filter bank) or processing time is sacrificed in order to "scan" through a range of parameter values. The following analysis, which assumes the use of as many filters as there are parameter values, therefore corresponds to the performance of an ideal receiver. It should also be pointed out that the system is ideal in other ways. For example, the mere assumption that the form of the signal is known presupposes perfect oscillator stability and exact knowledge of the characteristics of the system circuitry.

To facilitate the analysis, we assume that the filters are arranged so that they each have the same rms noise output in the absence of a signal. This is most easily done by so parametrizing the family of signals that

$$\int_{-\infty}^{\infty} | s(t; \alpha_1 \cdots \alpha_n) |^2 dt = 1 \tag{4}$$

for all $(\alpha_1 \cdots \alpha_n)$. This convention affects the physical meaning of the amplitude A. In fact, the total energy contained in the signal $Ae^{i\delta}s(t; \alpha_1 \cdots \alpha_n)$ is now simply $E_s = \frac{1}{2}A^2$ [our convention is that the average power association with a modulation function $z(t)$ is $\frac{1}{2}|z(t)|^2$]. Our detection criterion is now

$$\operatorname*{Max}_{\alpha_1 \cdots \alpha_n} \left| \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} z(t) dt \right| \geq \text{const.}, \tag{5}$$

and the set of values $(\hat{\alpha}_1 \cdots \hat{\alpha}_n)$, which provide this maximum, constitutes our estimation of the parameters. From now on we ignore the detection aspect of the problem, and concentrate on obtaining some properties of the estimates $(\hat{\alpha}_1 \cdots \hat{\alpha}_n)$.

Suppose that the actual signal present in the return has the modulation $A_0 e^{i\delta_0} s(t; \alpha_1^0 \cdots \alpha_n^0)$. Then the estimates must maximize the quantity

$$\left| A_0 e^{i\delta_0} \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} s(t; \alpha_1^0 \cdots \alpha_n^0) dt + \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} n(t) dt \right|,$$

which is the same as maximizing

$$\left| \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} s(t; \alpha_1^0 \cdots \alpha_n^0) dt + \frac{e^{-i\delta_0}}{A_0} \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} n(t) dt \right|.$$

The first term here is called the ambiguity function.

$$Q(\alpha_1 \cdots \alpha_n; \alpha_1^0 \cdots \alpha_n^0) = \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} s(t; \alpha_1^0 \cdots \alpha_n^0) dt. \tag{6}$$

If we put

$$X(\alpha_1 \cdots \alpha_n) = \frac{e^{-i\delta_0}}{A_0} \int_{-\infty}^{\infty} \overline{s(t; \alpha_1 \cdots \alpha_n)} n(t) dt, \tag{7}$$

then $(\hat{\alpha}_1 \cdots \hat{\alpha}_n)$ must maximize

$$| Q(\alpha_1 \cdots \alpha_n; \alpha_1^0 \cdots \alpha_n^0) + X(\alpha_1 \cdots \alpha_n) |. \tag{8}$$

The noise term, $X(\alpha_1 \cdots \alpha_n)$, is a complex Gaussian function of n variables, with zero mean, and generalized covariance

$$\begin{aligned} E\overline{X(\alpha_1 \cdots \alpha_n)} X(\alpha_1' \cdots \alpha_n') &= \frac{2N_0}{A_0^2} \int_{-\infty}^{\infty} \overline{s(t; \alpha_1' \cdots \alpha_n')} s(t; \alpha_1 \cdots \alpha_n) dt \\ &= \frac{N_0}{E_s} Q(\alpha_1' \cdots \alpha_n'; \alpha_1 \cdots \alpha_n). \end{aligned} \tag{9}$$

The estimation problem thus reduces to a location of the maximum of a multidimensional Gaussian process. This process is the absolute magnitude of the sum of a "signal" $Q(\alpha_1 \cdots \alpha_n; \alpha_1^0 \cdots \alpha_n^0)$, and a "noise" $X(\alpha_1 \cdots \alpha_n)$. Since this noise is completely characterized by the ambiguity function itself (together with the SNR), the ambiguity function defines the entire estimation problem.

According to our convention (4), we have

$$Q(\alpha_1 \cdots \alpha_n; \alpha_1 \cdots \alpha_n) = 1 \tag{10}$$

for all $(\alpha_1 \cdots \alpha_n)$. Further, by the Schwartz inequality,

$$| Q(\alpha_1' \cdots \alpha_n'; \alpha_1 \cdots \alpha_n) | \leq 1, \tag{11}$$

so that our "signal," $Q(\alpha_1 \cdots \alpha_n; \alpha_1^0 \cdots \alpha_n^0)$, always has a maximum at the true values $(\alpha_1^0 \cdots \alpha_n^0)$. This also shows that for very large values of the SNR, the true values of the parameters will always be among the possible estimates made by the matched-filter estimation system. Other values will also be among the estimates if the ambiguity function attains a value very near unity for incorrect values of the parameters. This, of course, is the ambiguity problem for single targets.

For large values of E_s/N_0, and in the absence of ambiguities, the estimates $(\hat{\alpha}_1 \cdots \hat{\alpha}_n)$ will be near the true values $(\alpha_1^0 \cdots \alpha_n^0)$. By expanding $Q(\alpha_1 \cdots \alpha_n; \alpha_1^0 \cdots \alpha_n^0)$, and $X(\alpha_1 \cdots \alpha_n)$ as functions of $(\alpha_1 \cdots \alpha_n)$ in Taylor series about the point $\alpha_k = \alpha_k^0$, and keeping only terms to second order, we can obtain first-order approximations to the estimates $(\hat{\alpha}_1 \cdots \hat{\alpha}_n)$. This procedure is carried out in detail,[2] and the result is that the errors $\hat{\alpha}_k - \alpha_k^0$ are Gaussian random variables with zero means (so that the estimates are unbiased) and

moment matrix

$$M_{kl} = E(\hat{\alpha}_k - \alpha_k{}^0)(\hat{\alpha}_l - \alpha_l{}^0), \tag{12}$$

whose inverse is related to the ambiguity function by the fundamental equation

$$(M^{-1})_{kl} = -\frac{A_0{}^2}{2N_0}\frac{\partial^2}{\partial\alpha_k\partial\alpha_l}\Big|\, Q(\alpha_1 \cdots \alpha_n;$$
$$\alpha_1{}^0 \cdots \alpha_n{}^0)\Big|^2\Big|_{(\alpha_1 \cdots \alpha_n)=(\alpha_1{}^0 \cdots \alpha_n{}^0)}. \tag{13}$$

The entire approximation procedure has a simple geometrical interpretation. The main peak of the ambiguity function [*i.e.*, the region near $(\alpha_1{}^0 \cdots \alpha_n{}^0)$] is approximated by a quadratic form in the differences $\alpha_k - \alpha_k{}^0$. This quadratic form (which corresponds to a family of ellipsoids in the parameter space) has a maximum value of unity at $(\alpha_1{}^0 \cdots \alpha_n{}^0)$ and matches the actual ambiguity function in curvature at this point. The characteristics of the approximate estimators are then determined entirely by the coefficients of curvature of the ambiguity function there. It is thus clear that a desirable ambiguity function has a sharp central peak and low secondary maxima.

Range, Velocity and Acceleration

If a signal of known form, say $s(t)$ (complex modulation), is transmitted in a monostatic radar, then the echo from a stationary target of small size (relative to the smallest modulation wavelength) will be

$$Ae^{i\delta}s(t-\tau)e^{-i\omega_0\tau}.$$

This represents the attenuation A, reflection phase shift δ, and range delay $\tau = 2r/c$, where r is the range. If the target moves, all these quantities will vary with time. Over a short period for a distant target, the attenuation can be assumed constant, but the delay associated with an arrival at time t will vary with t. We represent this delay by $\Delta(t)$, so that $c\Delta(t)$ is the distance travelled by a photon which arrives at time t. Now, the form of the returned signal modulation is

$$Ae^{i\delta}s(t-\Delta(t))e^{-i\omega_0\Delta(t)}. \tag{14}$$

Suppose the range to the target is known as a function of t, say $r = r(t)$. Then, since $c\Delta(t)$ is twice the distance to the point where the target was when reflection took place (for the photon which arrives at time t), and since this reflection took place $\Delta(t)/2$ seconds before arrival, we have the functional relation,

$$c\Delta(t) = 2r\left[t - \frac{\Delta(t)}{2}\right], \tag{15}$$

between range and delay. The general case of motion with constant acceleration in space leads to a very complicated dependence of $\Delta(t)$ on time. However, for relatively short times we can approximate $\Delta(t)$ by a quadratic function of t, and related coefficients to parameters of the motion. In particular, we choose the following form for $\Delta(t)$ in terms of three arbitrary parameters, τ, α, and β:

$$\Delta(t) = \tau - \frac{\beta}{\omega_0}(t-\tau) - \frac{\alpha}{\omega_0}(t-\tau)^2. \tag{16}$$

According to (16),

$$\begin{aligned}\Delta(\tau) &= \tau\\ \dot{\Delta}(\tau) &= -\frac{\beta}{\omega_0},\\ \ddot{\Delta}(\tau) &= -\frac{2\alpha}{\omega_0}\end{aligned} \tag{17}$$

where the dot stands for time-derivative. By evaluating identity (15) and its first two time derivatives at $t=\tau$, and using (17), we find

$$\begin{aligned}r(\tau/2) &= \frac{c\tau}{2}\\ \dot{r}(\tau/2) &= -\frac{\beta c}{\beta + 2\omega_0}\\ \ddot{r}(\tau/2) &= -\frac{8\alpha c\omega_0{}^2}{(\beta+2\omega_0)^3}.\end{aligned} \tag{18}$$

Thus τ has the significance of the arrival time of a photon which leaves the transmitter at $t=0$ (arriving at the target at $\tau/2$). The parameters β and α determine the range rate and range acceleration of the target at the time $(t=\tau/2)$ of reflection of this photon. Note that $\dot{r}(\tau/2)$ and $\ddot{r}(\tau/2)$ are approximately given by

$$\begin{aligned}\dot{r}(\tau/2) &\approx -\frac{\beta c}{2\omega_0}\\ \ddot{r}(\tau/2) &\approx -\frac{\alpha c}{\omega_0}.\end{aligned} \tag{19}$$

We shall not develop here the further relations connecting the derivatives of range with components of the vector velocity and acceleration.

If we substitute (16) into (14), we find that the general form assumed for the radar echo,

$$Ae^{i\delta}s\left[t - \tau + \frac{\beta}{\omega_0}(t-\tau) + \frac{\alpha}{\omega_0}(t-\tau)^2\right]e^{i\beta(t-\tau)+i\alpha(t-\tau)^2}, \tag{20}$$

in which the term $(-\omega_0\tau)$, the RF phase delay, has been lumped in with the reflection phase shift. Since the modulation generally consists of frequencies much lower than ω_0, the terms in α and β in the argument of $s(t)$ have relatively small effect. For the sake of simplicity, we shall drop these terms, although they could be important for very wide-band systems and hypervelocity targets. Thus, we begin our analysis with the following family of signals:

$$Ae^{i\delta}s(t-\tau)e^{i\alpha(t-\tau)^2+i\beta(t-\tau)}. \tag{21}$$

In accordance with our normalization convention, (4), we assume that

$$\int_{-\infty}^{\infty} |\, s(t)\,|^2 dt = 1. \tag{22}$$

The ambiguity function takes the form

$$\begin{aligned} &Q(\tau', \alpha', \beta'; \tau, \alpha, \beta) \\ &= \int_{-\infty}^{\infty} \overline{s(t-\tau')} s(t-\tau) e^{i\alpha(t-\tau)^2 + i\beta(t-\tau) - i\alpha'(t-\tau')^2 - i\beta'(t-\tau')} dt \\ &= \exp i[\alpha(\tau'-\tau)^2 + \beta(\tau'-\tau)] q(\alpha'-\alpha, \beta'-\beta, \tau'-\tau; \alpha), \end{aligned} \tag{23}$$

where we have put

$$q(\alpha, \beta, \tau; \alpha^0) = \int_{-\infty}^{\infty} \overline{s(t)} s(t+\tau) e^{-i\alpha t^2 - i(\tau - 2\alpha^0\tau)t} dt. \tag{24}$$

The dependence of the ambiguity function on the differences $\tau'-\tau$, $\alpha'-\alpha$, and $\beta'-\beta$ is supplemented by an explicit dependence on α. This will have, as a consequence, a dependence of the estimation errors on the true acceleration.

The phase factor by which Q differs from q in (23) can have no effect on the small-error moment matrix M_{kl} given by (12) and (13), since this matrix depends only on $|Q|^2 = |q|^2$. Also, referring to the general equation (8), the statistical properties of $X(\alpha_1 \cdots \alpha_n)$ are unchanged by such a phase factor. Therefore q itself, as given by (24), can serve as the ambiguity function for our problem.

Various general properties of the ambiguity function can be easily obtained from (24). For example, the integral of $|q|^2$ over the time-Doppler plane, for fixed values of α and α^0 is constant, and equal to 2π. Special formulas for q along the various coordinate axes are also easily obtained. In the special case of a periodic train of very short pulses, (24) provides a means of studying the ambiguities which arise. We find that corresponding to the "blind speed," $v_b = c/2f_0T$ (where T is the pulse repetition period), there is also a "blind acceleration," $a_b = c/2f_0T^2$. In other words, accelerations measured with this waveform will be ambiguous to the extent of an unknown multiple of a_b.

The Error Moment Matrix

In order to apply (12) and (13) for the error moment matrix in the strong-signal case, let us agree to arrange the parameters as follows: $\alpha_1 = \alpha$, $\alpha_2 = \beta$ and $\alpha_3 = \tau$. Then, using (23), we have

$$\begin{aligned} (M^{-1})_{kl} &= -\frac{A_0^2}{2N_0} \frac{\partial^2}{\partial\alpha_k \partial\alpha_l} \cdot \Big| q(\alpha_1 - \alpha_1^0, \alpha_2 - \alpha_2^0, \alpha_3 - \alpha_3^0; \alpha_1^0) \Big|^2_{\alpha=\alpha^0} \\ &= -\frac{A_0^2}{2N_0} \frac{\partial^2}{\partial\alpha_k \partial\alpha_l} \Big| q(\alpha_1\alpha_2\alpha_3; \alpha_1^0) \Big|^2_{\alpha_k=0} \\ &= \frac{A_0^2}{N_0} \operatorname{Re} \left\{ \frac{\partial^2}{\partial\alpha_k \partial\alpha_l} q(\alpha_1\alpha_2\alpha_3; \alpha_1^0) \Big|_{\alpha_k=0} \right. \\ &\quad \left. + \frac{\partial q(\overline{\alpha_1\alpha_2\alpha_3}; \alpha_1^0)}{\partial\alpha_k} \Big|_{\alpha=0} \frac{\partial q(\alpha_1\alpha_2\alpha_3; \alpha_1^0)}{\partial\alpha_l} \Big|_{\alpha=0} \right\}. \end{aligned} \tag{25}$$

To get the last form in (25) we have used the fact that $q(0, 0, 0; \alpha_1^0) = 1$. The actual computation of the matrix elements of M^{-1}, carried out by substituting (24) into (25), is quite tedious, and we simply quote the result. To avoid lengthy writing, we put

$$s(t) = u(t) \exp\left[i \int_0^t \omega(t')dt' \right], \tag{26}$$

where $u(t)$ is the (real) amplitude modulation and $\omega(t)$ is the (real) frequency modulation and we use the following abbreviations:

$$m_2 = \int_{-\infty}^{\infty} t^2 u(t)^2 dt - \left[\int_{-\infty}^{\infty} t u(t)^2 dt \right]^2, \tag{27}$$

$$m_3 = \int_{-\infty}^{\infty} t^3 u(t)^2 dt - \int_{-\infty}^{\infty} t u(t)^2 dt \int_{-\infty}^{\infty} t^2 u(t)^2 dt, \tag{28}$$

$$m_4 = \int_{-\infty}^{\infty} t^4 u(t)^2 dt - \left[\int_{-\infty}^{\infty} t^2 u(t)^2 dt \right]^2, \tag{29}$$

$$\delta_1 = \int_{-\infty}^{\infty} t\omega(t) u(t)^2 dt - \int_{-\infty}^{\infty} t u(t)^2 dt \int_{-\infty}^{\infty} \omega(t) u(t)^2 dt, \tag{30}$$

$$\delta_2 = \int_{-\infty}^{\infty} t^2 \omega(t) u(t)^2 dt - \int_{-\infty}^{\infty} t^2 u(t)^2 dt \int_{-\infty}^{\infty} \omega(t) u(t)^2 dt, \tag{31}$$

$$\omega_2^2 = 4\pi^2 \int_{-\infty}^{\infty} f^2 |\, g(f)\,|^2 df - \left[2\pi \int_{-\infty}^{\infty} f |\, g(f)\,|^2 df \right]^2. \tag{32}$$

These quantities have the character of moments of the (normalized) distribution $u(t)^2$. The last expression has the form of a moment over the Fourier transform,

$$g(f) = \int_{-\infty}^{\infty} s(t) e^{-2\pi i f t} dt. \tag{33}$$

In terms of these quantities we have, for M^{-1},

$$M^{-1} = \frac{A_0^2}{N_0} \cdot \begin{bmatrix} m_4 & m_3 & -\delta_2 - 2\alpha^0 m_3 \\ m_3 & m_2 & -\delta_1 - 2\alpha^0 m_2 \\ -\delta_2 - 2\alpha^0 m_3 & -\delta_1 - 2\alpha^0 m_2 & \omega_2^2 + 4\alpha^0\delta_1 + 4(\alpha^0)^2 m_2 \end{bmatrix}. \tag{34}$$

The determinant of this matrix reduces considerably. In fact, it is independent of α^0:

$$\begin{aligned} \operatorname{Det}(M^{-1}) &= \left(\frac{A_0^2}{N_0} \right)^3 \{ (m_4 m_2 - m_3^2)\omega_2^2 - m_2\delta_1^2 \\ &\quad + 2m_3\delta_1\delta_2 - m_2\delta_2^2 \} \equiv \frac{A_0^2}{N_0} \Delta. \end{aligned} \tag{35}$$

It is also easily shown that this determinant is invariant to the choice of time origin.

Other useful properties are the facts that Δ is non-negative and insensitive to FM, up to quadratic, that may be present in the waveform. The first property will be proved later; here we prove the second. If

$$\omega(t) = a + bt + ct^2, \tag{36}$$

which represents a general combination of linear and quadratic FM (the constant term simply changes the carrier frequency), we find that

$$\begin{aligned} \delta_1 &= bm_2 + cm_3, \\ \delta_2 &= bm_3 + cm_4. \end{aligned} \tag{37}$$

It can also be shown that, in general,

$$\begin{aligned} \omega_2^2 = &\int_{-\infty}^{\infty} \left| \omega(t)^2 u(t)^2 + \dot{u}(t)^2 \right| dt \\ &- \left[\int_{-\infty}^{\infty} \omega(t) u(t)^2 dt \right]^2, \end{aligned} \tag{38}$$

which reduces to

$$b^2 m_2 + 2bcm_3 + c^2 m_4 + \int_{-\infty}^{\infty} [\dot{u}(t)]^2 dt, \tag{39}$$

in the case described by (36). Substitution of (37) and (39) into (35) then results in

$$\Delta = (m_4 m_2 - m_3^2) \int_{-\infty}^{\infty} [\dot{u}(t)]^2 dt, \tag{40}$$

which is completely insensitive to the presence of the FM.

Turning now to the moment matrix itself, we find that the cofactor of the (11)-element in the matrix of (34) is simply $m_2\omega_2^2 - \delta_1^2$, which yields

$$\begin{aligned} M_{11} &= \frac{N_0}{A_0^2} \frac{m_2\omega_2^2 - \delta_1^2}{\Delta} \\ &= \frac{N_0}{A_0^2} \left[\left(m_4 - \frac{m_3^2}{m_2} \right) - \frac{(m_3\delta_1 - m_2\delta_2)^2}{m_2(m_2\omega_2^2 - \delta_1^2)} \right]^{-1}. \end{aligned} \tag{41}$$

This element, which determines the variance of the acceleration error, is independent of α^0 and is invariant to the choice of time origin. In the special case (36), we find that

$$m_2\omega_2^2 - \delta_1^2 = m_2 \int_{-\infty}^{\infty} [\dot{u}(t)]^2 dt + c^2(m_4 m_2 - m_3^2), \tag{42}$$

so that the acceleration accuracy is affected only by the quadratic component of the FM. From the second form of (41) it is clear that any form of FM can only degrade the acceleration accuracy, since

$$\begin{aligned} m_4 m_2 - m_3^2 &\geq 0 \\ m_2\omega_2^2 - \delta_1^2 &\geq 0, \end{aligned} \tag{43}$$

as can readily be shown by the use of the Schwartz inequality.

If there is no phase modulation, M_{11} reduces to

$$M_{11} = \frac{N_0}{A_0^2} \left(m_4 - \frac{m_3^2}{m_2} \right)^{-1}. \tag{44}$$

Now M_{11} is the variance of the error in α, and α is related to acceleration by relation (19), so that the standard deviation Δa of the acceleration error is given by

$$\Delta a = \frac{\lambda_0}{2\pi} (M_{11})^{1/2} = \frac{\lambda_0}{2\pi \sqrt{\frac{2E_s}{N_0} \left(m_4 - \frac{m_3^2}{m_2} \right)}}. \tag{45}$$

Suppose that the envelope $u(t)$ vanishes outside an interval of time of duration T_0, but is otherwise arbitrary. Then it can be shown, that among this class of functions, the one which maximizes the combination $(m_4 - m_3^2/m_2)$ consists of three very short pulses—two of equal height at the ends of the interval, and one at the center of the interval which contains twice the energy of either of the other pulses. For this waveform,

$$m_4 - \frac{m_3^2}{m_2} = T_0^4/64, \tag{46}$$

and the acceleration accuracy is

$$\Delta a = \frac{4\lambda_0}{\pi \sqrt{\frac{2E_s}{N_0}} T_0^2}.$$

In general, for waveforms of duration T_0, we write

$$m_4 - \frac{m_3^2}{m_2} = f_a T_0^4/64, \tag{47}$$

where $f_a \leq 1$ is an "acceleration form-factor." Then

$$\Delta a = \frac{4\lambda_0}{\pi \sqrt{\frac{2E_s}{N_0}} f_a T_0^2}. \tag{48}$$

In evaluating these form factors for symmetrical waveforms, it is convenient to choose the time origin at the midpoint of the time interval so that $m_3 = 0$. For example, if $[u(t)]^2$ consists of three pulses of equal energy, placed at the ends and center of the interval, we find $f_a = 8/9 = 89$ per cent. For one long pulse of constant height and length T_0, we find $f_a = 64/180 = 35.5$ per cent. In this latter case, (48) specializes to

$$\Delta a = \frac{3\sqrt{5}\,\lambda_0}{\pi \sqrt{\frac{2E_s}{N_0}} T_0^2}. \tag{49}$$

It can also be shown that (49) holds for a train of many short pulses, equi-spaced over a time interval of length T_0.

The formulas for the variances of the range and velocity errors, and their correlation coefficient, are quite complicated in the general case. Even with the same waveform, the variances of these errors are different if acceleration is present and must be concurrently measured, than if the acceleration is known or is absent.[5] Some of the terms in the velocity error even depend on the true value of the unknown acceleration. To facilitate the exposition, we shall first develop the formulas for the familiar case when acceleration is known to be absent, since this situation has many analogies to the general case.

The ambiguity function for range and velocity only is obtained from (24) by putting $\alpha=\alpha^0=0$. The inverse of the moment matrix is then the relevant part of (34) (with $\alpha^0=0$).

We have, namely,

$$M^{-1}=\frac{A_0^2}{N_0}\begin{bmatrix} m_2 & -\delta_1 \\ -\delta_1 & \omega_2^2 \end{bmatrix}; \tag{50}$$

hence,

$$M=\frac{N_0/A_0^2}{m_2\omega_2^2-\delta_1^2}\begin{bmatrix} \omega_2^2 & \delta_1 \\ \delta_1 & m_2 \end{bmatrix}. \tag{51}$$

The (11)-element refers to velocity and (22)-element refers to range. The standard deviations of the range and velocity errors are

$$\Delta R=\frac{c\Delta\tau}{2}=\frac{c}{2\sqrt{\frac{2E_s}{N_0}}\sqrt{\omega_2^2-\frac{\delta_1^2}{m_2}}}, \tag{52}$$

$$\Delta v=\frac{\lambda_0\Delta\beta}{4\pi}=\frac{\lambda_0}{4\pi\sqrt{\frac{2E_s}{N_0}}\sqrt{m_2-\frac{\delta_1^2}{\omega_2^2}}}. \tag{53}$$

It should be observed that these error formulas are different from those which would be valid if range or velocity were being measured singly and not together. The range accuracy attainable if only range is measured is given by (53) with $\delta_1=0$. Similarly, the velocity error, if only velocity is measured, is given by (52) with $\delta_1=0$. It is interesting that the errors are never smaller when both quantities are measured concurrently. The quantity δ_1 which modifies these errors also determines the correlation between them; the normalized correlation coefficient for the range and Doppler errors is just $\delta_1/\sqrt{m_2\omega_2^2}$. From its definition (30), it is clear that odd-power frequency modulation (together with a symmetric envelope function) will yield nonvanishing δ_1. This correlation is well known for the case of linear frequency modulation; if we put $\omega(t)=at$, we find that the normalized correlation coefficient becomes equal to

$$\left\{1+\frac{1}{m_2a^2}\int_{-\infty}^{\infty}[\dot{u}(t)]^2dt\right\}^{-1/2}, \tag{54}$$

which can be nearly unity for large a.

The best velocity accuracy is attained for those waveforms that yield $\delta_1=0$, and hence coincides with the absence of correlation (*e.g.*, no frequency modulation or even-power frequency modulation). In this case the velocity error is

$$\Delta v=\frac{\lambda_0}{4\pi\sqrt{\frac{2E_s}{N_0}m_2}}. \tag{55}$$

It is easily seen that the best waveform for velocity measurement, among those waveforms of duration T_0, is one consisting of two equal short pulses at the ends of the interval. In this case we get $m_2=T_0^2/4$. In general for such waveforms, we put $m_2=f_vT_0^2/4$, where $f_v\leq 1$ is a velocity form factor, so that

$$\Delta v=\frac{\lambda_0}{2\pi\sqrt{\frac{2E_s}{N_0}f_vT_0}}. \tag{56}$$

If the waveform consists of a long flat pulse of length T_0, we find that $f_v=\frac{1}{3}$, and

$$\Delta v=\frac{\sqrt{3}\,\lambda_0}{2\pi\sqrt{\frac{2E_s}{N_0}T_0}}. \tag{57}$$

For a train of N pulses, each of length τ, spanning a total time T_0, we find that

$$f_v=\frac{1}{3}\,\frac{N-1}{N+1}\left(1-2\,\frac{\tau}{T_0}+\frac{2N}{N+1}\,\frac{\tau^2}{T_0^2}\right), \tag{58}$$

an expression which reduces to 1/3 when $T_0=N\tau$ (one long pulse) and when $\tau=0$, $N\to\infty$ (a train of many infinitely short pulses).

Regarding the range accuracy, we note that FM tends to increase the error by way of the δ_1-term in (52). However, ω_2^2 is also sensitive to FM. Using the form (38), we have

$$\omega_2^2-\frac{\delta_1^2}{m_2}=\int_{-\infty}^{\infty}[\omega(t)^2u(t)^2+\dot{u}(t)^2]dt-\left[\int_{-\infty}^{\infty}\omega(t)u(t)^2dt\right]^2-\frac{\delta_1^2}{m_2}. \tag{59}$$

For a general combination of linear and quadratic FM, namely, $\omega(t)=a+bt+ct_2$, as in (36), we note (37) and (39) and find that

$$\omega_2^2-\frac{\delta_1^2}{m_2}=\int_{-\infty}^{\infty}[\dot{u}(t)]^2dt+c^2\left(m_4-\frac{m_3^2}{m_2}\right). \tag{60}$$

[5] A similar situation is the dependence of the detection probability curve on the number of parameters which are being concurrently estimated. See (2).

Therefore, linear FM has no effect on the range accuracy, while quadratic FM improves the attainable accuracy. In the absence of FM, we have

$$\Delta R = \frac{c}{2\omega_2\sqrt{\frac{2E_s}{N_0}}}, \tag{61}$$

and ω_2^2 can be evaluated from (32). If the signal spectrum vanishes outside a band of width B_0, then $\omega_2^2 \leq \pi^2 B_0^2$, by an argument similar to that used in the time domain above. Equality is attained for a signal spectrum consisting of two spikes at the ends of the band. In general, for band-limited waveforms, we put $\omega_2^2 = f_R \pi^2 B_0^2$, which defines a range form factor, and leads to the formula

$$\Delta R = \frac{c}{2\pi\sqrt{\frac{2E_s}{N_0} f_R B_0}}. \tag{62}$$

Let us now develop the velocity- and range-error formulas for the general case of unknown acceleration. After some computation one finds that

$$M_{22} = \frac{N_0}{A_0^2} \frac{m_4\omega_2^2 - \delta_2^2 + 4\alpha^0(m_4\delta_1 - m_3\delta_2) + 4(\alpha^0)^2(m_4 m_2 - m_3^2)}{(m_4 m_2 - m_3^2)\omega_2^2 - m_4\delta_1^2 + 2m_3\delta_1\delta_2 - m_2\delta_2^2}, \tag{63}$$

an unwieldy expression which depends on the true acceleration and is not invariant to a change of time origin. In other words, for an accelerating target, the quality of the velocity measurement depends upon the time at which it is measured. In general, the optimum time (which choice also removes the velocity-acceleration correlation) depends upon α^0, which is unknown in advance. It is hard to see what the general effect of FM is in this case; however, if FM is absent, (63) simplifies to

$$M_{22} = \frac{N_0}{A_0^2}\left\{\frac{1}{m_2 - \frac{m_3^2}{m_4}} \frac{4(\alpha^0)^2}{\omega_2^2}\right\}. \tag{64}$$

The corresponding velocity error is

$$\Delta v = \frac{\lambda_0}{4\pi\sqrt{\frac{2E_s}{N_0}}}\sqrt{\frac{1}{m_2 - \frac{m_3^2}{m_4}} + \frac{1}{f_R}\left(\frac{4a^0}{\lambda_0 B_0}\right)^2}, \tag{65}$$

where a^0 is the true acceleration and we have put $\omega_2^2 = f_R\pi^2 B_0^2$, as before. For many problems, the acceleration-dependent term will be negligible. For example, if a^0 is the acceleration due to gravity, $\lambda_0 = 1$ foot, and $B_0 = 10$ Mc, we find $\lambda_0 B_0/4a^0 = 78{,}000$ seconds, which will certainly make a negligible contribution to (65) for most observation times of interest.

If we neglect this term, we have

$$\Delta v = \frac{\lambda_0}{4\pi\sqrt{\frac{2E_s}{N_0}}\sqrt{m_2 - \frac{m_3^2}{m_4}}}, \tag{66}$$

which is still not the same as in the case with no acceleration. It is easily seen that Δv is smallest if the time origin is shifted to the point $t_0 = -m_3/2m_2$, which is equivalent to displacing the waveform to make $m_3 = 0$. Note that this choice can be made *à priori*, since it does not involve α^0. The velocity error now reduces to the previous value, given by (55) and its special cases, (56) and (57). It can also be shown that the velocity-acceleration correlation is very small for this choice of time reference.

The range error is, in general, given by

$$\Delta R = \frac{c}{2}(M_{33})^{1/2} = \frac{c}{2\sqrt{\frac{2E_s}{N_0}}}\left[\omega_2^2 - \frac{m_4\delta_1^2 - 2m_3\delta_1\delta_2 + m_2\delta_2^2}{m_4 m_2 - m_3^2}\right]^{-1/2}. \tag{67}$$

We note that the determinant Δ can be expressed as

$$\Delta = \frac{N_0}{A_0^2}\frac{m_4 m_2 - m_3^2}{M_{33}}, \tag{68}$$

which is non-negative, as was stated above. It can be shown that for any combination of linear and quadratic FM, (67) is unchanged; it reduces to

$$\Delta R = \frac{c}{2\sqrt{\frac{2E_s}{N_0}}}\left\{\int_{-\infty}^{\infty}[\dot{u}(t)]^2 dt\right\}^{-1/2}. \tag{69}$$

Only by using cubic or higher types of FM can the attainable range accuracy be increased. Of course, FM is used in pulse compression schemes, as a rule, to reduce the sidelobes of the ambiguity function in range, and not to improve range accuracy itself. In the absence of cubic or higher FM, (67) and (69) coincide with (61) and its specialization (62).

The normalized correlation coefficients, ρ_{ij}, can also be obtained, although the general expressions are very complicated. If the signal carries no FM, and if the acceleration is small enough so that we may neglect terms dependent on α^0, we find that ρ_{13} and ρ_{23} vanish, while ρ_{12} reduces to $(-m_3/m_2)$, which can be made to vanish by proper choice of the time origin.

Masers for Radar Systems Applications*

H. R. SENF†, SENIOR MEMBER, IRE, F. E. GOODWIN†, MEMBER, IRE,
J. E. KIEFER†, AND K. W. COWANS‡

Summary—This paper is intended to provide the systems engineer with a practical introduction to the use of solid-state maser amplifiers in radars. Various environmental problems involved in the successful application of masers are discussed. An elementary survey of reflection-cavity and traveling-wave masers, together with some experimental results are presented. Another section treats the problem of saturation in masers and discusses some of the methods available for protecting masers from the TR leak-through pulses in radar. Progress made in the development of open- and closed-cycle liquid helium cryogenic systems suitable for masers is described. The authors' personal evaluations of the state of the art of ruby masers and closed-cycle helium refrigerators are given in appendixes.

Introduction

THE PROBLEM of reducing receiver noise has fascinated radar engineers since the earliest days of radar. The value of a reduction in receiver noise, in terms of improvement in detection range, becomes most dramatically apparent when one considers the increase in transmitter power required to achieve a comparable improvement. The results obtained by the combination of a few improvements in the circuits and components used in the front end of the receiver at little or no cost in size, weight, and power might require many hundreds of pounds of additional transmitter equipment.

The search for the ideal radar receiver has progressed through the years with outstanding contributions by many workers, but the end result always fell short of the goal. The maser, meaning "Microwave Amplification by Stimulated Emission of Radiation," has for the first time brought the goal within reach. The results achieved during the past few years with solid-state masers show that the detection range of a radar no longer need be significantly limited by the noise generated within the receiver.

With the advent of the maser we are tempted to speculate that the long quest for the ideal low-noise receiver is approaching an end. It will be shown in the following sections, however, that present-day masers leave many problems for future workers. This is particularly true of important secondary objectives such as simplicity of equipment and flexibility of performance.

System Considerations

The noise in a radar receiver with which the signal must compete in order to be detected is the sum of two parts: that which is generated within the receiver, and that which originates in sources external to the receiver. (In this paper "external" noise is defined as the total noise at the output of the antenna; thus it includes both the "sky" noise as defined by De Grasse, *et al.*,[1] and the integrated atmospheric and earth noise due to the secondary pattern of the antenna.) Throughout the early history of radar, the magnitude of the internal noise was so large in comparison with the external noise that the latter could be neglected without significantly changing the performance of the system. Prior to the advent of masers and parametric amplifiers, the noise temperature of a typical microwave receiver was about 1500°K. The external noise for most radar systems was between 50° and 300° K. Recently, as the result of special care taken in the design of the antenna, the external noise of a *C*-band ground receiving system looking at elevation angles greater than 30 degrees above the horizon in good weather has been reduced to less than 10°K.[1] Further measurements and analyses have shown that external noise for similar antennas and environmental conditions should be less than about 20°K over the frequency band of 1 to 10 kMc (see Fig. 1).[2] For receiving antennas operating outside the earth's atmosphere and not pointing directly toward the earth, sun, or moon, it is expected that external noise will be less than 10°K, throughout the radio frequency range above 1 kMc.[2] Simple calculations based on the radar-range equation show that the detection range of radar for space applications can be increased by as much as three times if the noise generated within the receiver can be reduced so that it is equal to the external noise. The solid-state maser amplifier integrated with a carefully designed antenna has demonstrated the feasibility of achieving this result.

Because of the size, weight, complexity, and operating problems inherent to early maser amplifiers, they have been slow to find applications in practical electronic systems. A maser, to be fully effective, must be placed at or near the feed point of the antenna (see Fig. 2). This arrangement precludes the introduction of internal noise by such components as rotary joints, waveguide runs, switches, etc. Each 0.1 db of loss in such devices placed ahead of the maser will contribute 7°K to the total internal receiver noise, assuming the lossy devices

* Received by the PGMIL, January 17, 1961.
† Hughes Res. Labs., Hughes Aircraft Co., Malibu, Calif.
‡ Aerospace Engrg. Div., Hughes Aircraft Co., Culver City, Calif.

[1] R. W. De Grasse, D. C. Hogg, E. A. Ohm, and H. E. D. Scovil, "Ultra-low and noise measurements using a horn reflector antenna and a traveling-wave maser," *J. Appl. Phys.*, vol. 30, p. 2013; December, 1959.

[2] R. L. Forward and F. Richey, "Effects of external noise on radar performance," *Microwave J.*, vol. 3, pp. 73–80; December, 1960.

are at normal ambient temperatures. The antennas under consideration usually are large dishes mounted on high pedestals with feed horns suspended in positions that are awkward for mounting complex electronic and cryogenic apparatus.

Most of the masers that have been developed to date use rather large double dewars containing liquid nitrogen and liquid helium for the refrigeration system. Typically, these dewars must be refilled with both liquids at least once each day. The logistic problem of supplying the cryogenic fluids at the antenna site, plus the operational problem of transferring the fluids into the maser dewar, have presented two of the most serious obstacles to the use of masers.

Another difficulty with early masers has been the requirement for a large magnet to provide the dc field needed to tune the maser to the desired frequency. The magnet was placed outside the dewar, leading to an "air" gap in the magnetic circuit many times the dimension of the maser material. This gap caused the size and weight of the magnet to be very large.

Together, these problems until recently have limited the application of masers to a few experimental scientific measurements such as radio and radar astronomy.

During the last year or two, important progress has been made toward overcoming many of these obstacles. Several masers using small permanent magnets inside the dewar have been developed. A more compact open helium dewar-maser unit which eliminates the need for liquid nitrogen also has been developed at Hughes. For an X-band receiver, the maser equipment required to be mounted at the feed point occupies only about $\frac{1}{3}$ of a cubic foot, and weighs about 20 lbs. An artist's sketch of an early model of this unit is shown in Fig. 3.

The first models of closed-helium refrigerators for masers for ground systems applications have been developed. These equipments are, as yet, rather large but indicate promise for the use of practical refrigerators for ground applications in the near future.

A compact and relatively efficient closed-helium refrigerator for air- and missile-borne applications is in an advanced state of development. An airborne X-band maser in which the refrigerator and dewar units together will occupy about 1.3 cubic feet and weigh about 60 lbs. is under development at Hughes (see Fig. 4). About $1\frac{1}{2}$ kw of primary power is required to operate the refrigerator for a typical maser.

Until now (and probably for some time in the future), all masers have been custom designed and developed by

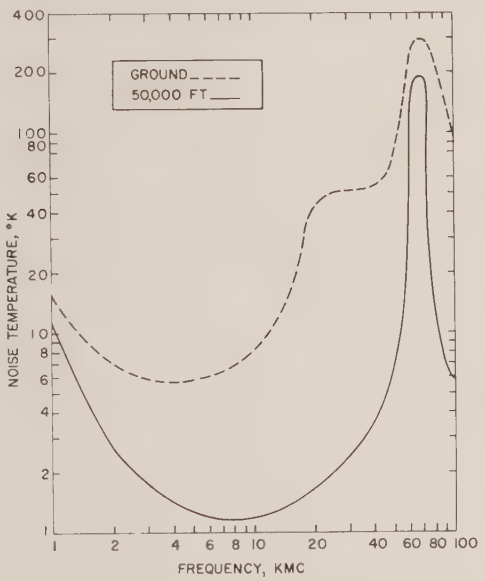


Fig. 1—External noise for ground and airborne antennas (beam 30° above horizon).

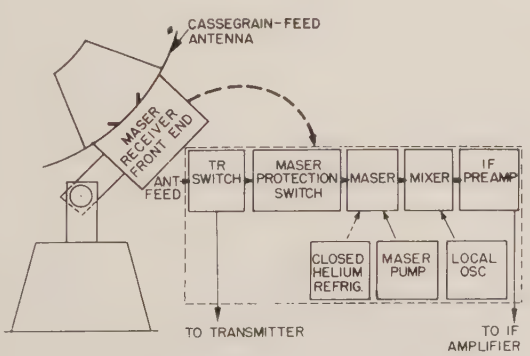


Fig. 2—Typical maser installation for ground radar.

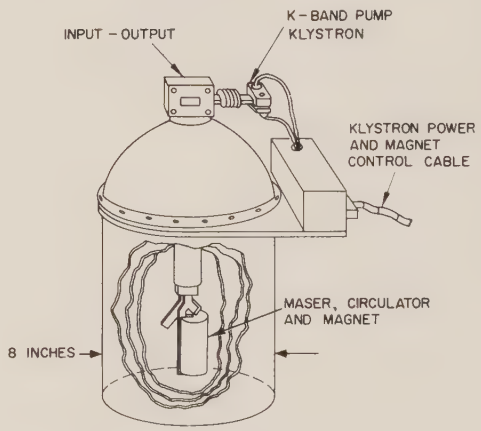


Fig. 3—X-band maser in compact open-helium dewar.

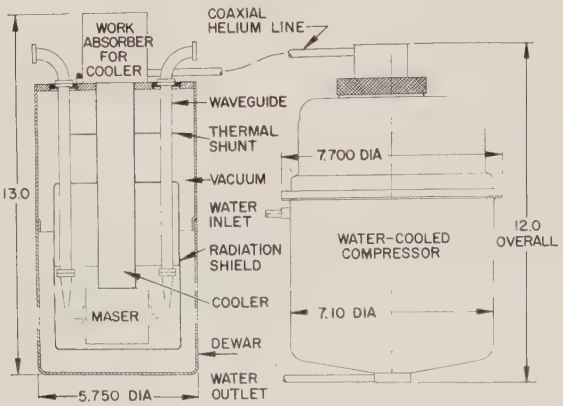


Fig. 4—Maser with integral closed-helium refrigerator.

research engineers. In general, only one of a kind has been made. The same is true of closed-helium refrigerators, and, to a lesser extent, open-helium dewars. Thus, the cost of prototype or experimental masers with closed refrigerators is still very high. The cost and complexity have been justified in the past for a number of critical measurements and specialized applications. During the next few years, as maser techniques are reduced to engineering technology, masers should become practical for an increasing range of system applications.

The authors' personal evaluation of the state of the art of ruby masers and closed-cycle helium refrigerators are given in Appendixes I and II.

Elements of Maser Technology

The problem of developing a maser technology suitable for a variety of systems applications consists of establishing the feasibility of operating over a wide range of frequencies with the necessary gain, bandwidth, tuning range, and stability. All of this must be done with equipment which can be operated continuously without the difficulties of filling the cryogenic system. Another basic element in achieving this capability is that of providing suitable maser materials for the various frequency ranges. Ruby, which is an aluminum oxide crystal with a small percentage doping of chromium ions, has proved to be successful for operation in the frequency range from a few hundred megacycles to 15 kMc. It has excellent physical properties and can withstand the shock of repeated temperature cyclings without damage.

In an effort to provide a comparable material suitable for the frequency range from 10 kMc to approximately 60 kMc, synthetic emerald with several values of chromium concentration is being investigated. Studies of the millimeter-wave spectroscopy of emerald have shown that it has excellent promise for this frequency region.[3] The microwave spectrum of emerald for the symmetric mode is illustrated in Fig. 5. This figure shows the manner in which the magnitude of the energy-level transitions between the four basic energy states of the chromium ion in emerald vary with an externally-applied dc magnetic field. The mode of operation shown is that of a broad-band maser amplifier at 35 kMc. This mode is particularly desirable because operation over a band of more than 1 kMc appears to be feasible without change of pump frequency.

It is not within the scope of this paper to discuss the internal physics of maser materials. It should be pointed out, however, that this subject is vital to the development of a suitable maser technology. The future maser engineer must be thoroughly acquainted with many technical problems previously considered to be within the realm of the solid-state physicist.

[3] J. E. Geusic, M. Peter, and E. O. Schulz-Du Bois, "Paramagnetic resonance spectrum of Cr^{+++} in emerald," *Bull. Am. Phys. Soc.*, ser. II, vol. 4, p. 21; January 28, 1959.

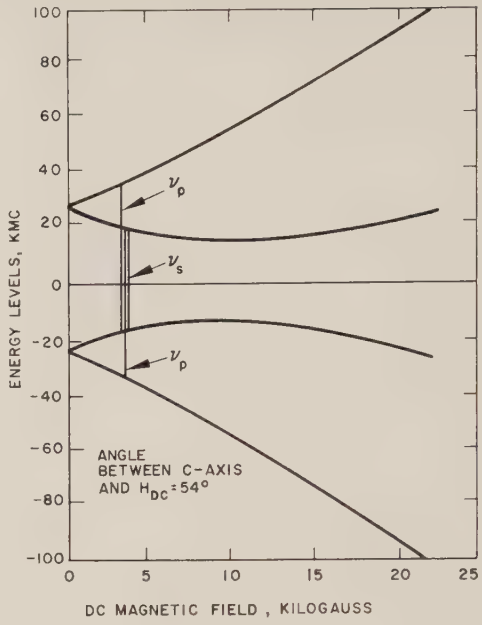


Fig. 5—Microwave spectra of emerald (symmetric model).

A typical maser amplifier consists of a microwave resonator which contains a piece of crystalline maser material. The circuit is resonant both at a signal frequency and at a so-called "pump" frequency. The circuit is cooled to a very low temperature, usually within a few degrees of absolute zero, and placed in a dc magnetic field. A few milliwatts of power are continuously injected into the circuit from the pump oscillator. The maser crystal absorbs and stores energy from the pump. When a signal is applied, it stimulates the emission of some of the stored energy. This energy is released as discrete quanta at signal frequency, each input quantum triggering the emission of many stored quanta. Amplification is obtained by the addition of the emitted energy to the input signal.

Reflection-Cavity Masers

Early masers were constructed in a form usually referred to as a reflection-cavity maser. Two variations of the reflection-cavity maser are shown in Fig. 6. This kind of maser requires a circulator for its operation, since the input and output signals from the maser resonator are present in the same transmission line. The figure on the left side of Fig. 6 describes the basic reflection-cavity maser configuration. Early *X*-band masers of this type exhibited a typical performance of 26-db gain with a 5-Mc bandwidth and a noise temperature of about 25°K.[4] The circulator was operated at room temperature and contributed about 15°K of the noise.[5]

[4] T. H. Maiman, "Solid-State Maser Amplifier X-Band," Hughes Res. Lab., Malibu, Calif., Final Progress Rept. on Signal Corps Contract DA 36-039SC-74951; November, 1958.

[5] A comparable maser which used a circulator cooled to liquid helium temperature had a measured noise temperature of 10°K.

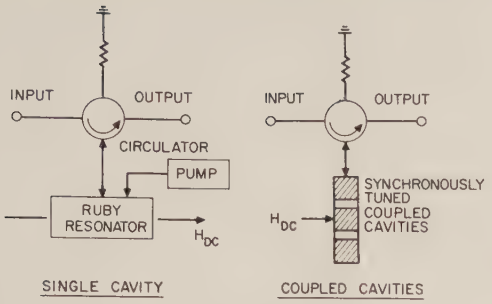


Fig. 6—Reflection-cavity masers.

The figure on the right side of Fig. 6 illustrates a method of broadbanding reflection masers. In this case, the single resonator is replaced by a system of coupled resonators. The amount of improvement which can be obtained in the gain-bandwidth product by this method varies with the absolute gain of the system. For a gain of 26 db, a two-resonator reflection maser has approximately $3\frac{1}{2}$ times the gain-bandwidth of a single-resonator maser.[6] This design is suitable for fixed-frequency systems requiring more gain-bandwidth product than can be achieved in the single-cavity maser.

An important maser engineering problem which already has been solved for some applications is that of replacing the large external electromagnet used in all early research on masers by a small permanent magnet located inside the dewar. Fig. 7 shows an *X*-band maser-dewar unit in which the permanent magnet is placed inside the dewar. This unit was developed at the Hughes Research Laboratories under Signal Corps sponsorship and is one of the first packaged maser units of this type to be delivered to a customer.[5] Fig. 8 shows a close-up of the resonator and the permanent magnet; these are mounted at the bottom of the dewar in liquid helium. This unit combines the advantages of the internal magnet, push-pull pumping, and 100 per cent filling factor. With a low-loss circulator placed outside the dewar at room temperature, the noise temperature of this unit was 25°K and the gain-bandwidth product was 105 Mc. This represented a significant advance for this type of maser at the time it was developed in 1958.

Fig. 9 shows the microwave circuitry of an *X*-band maser with a helium-cooled circulator. This circulator is of the polarization-rotation type. It also can be used to switch the input from the maser to a matched load by reversing the direction of the dc magnetic field. In this way the maser can be protected from the saturating effects of a local pulse signal, such as that obtained in a radar. The noise temperature of this maser amplifier was measured as 10°K. Essentially no noise was contributed by the circulator because of its operating temperature of 4.2°K. Subsequently, miniature cooled *Y*-type circulators that have improved bandwidth

[6] F. E. Goodwin, "One Port and Traveling-Wave Maser Using Coupled Cavities," Hughes Res. Labs., Malibu, Calif., Res. Rept. No. 173; November, 1960.

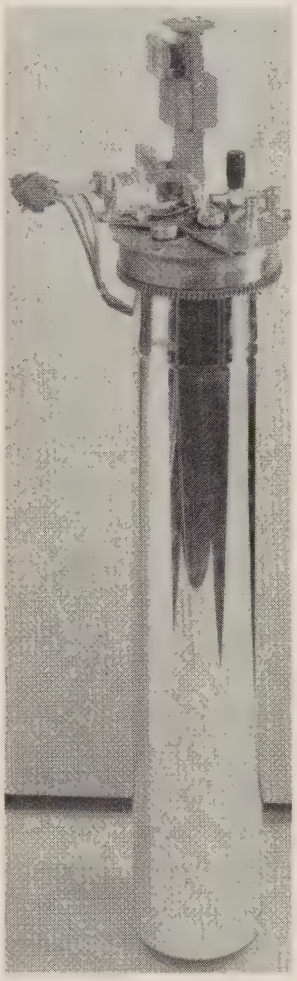


Fig. 7—*X*-band maser-dewar unit.

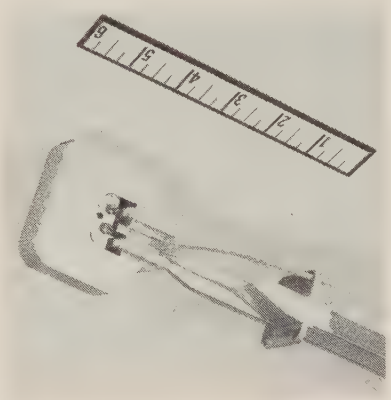


Fig. 8—*X*-band maser resonator and magnet.

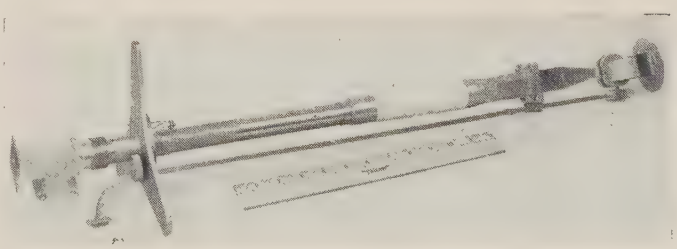


Fig. 9—Maser with cooled circulator.

characteristics and equally low noise have been developed. These will be valuable in making future masers more compact and efficient.

Traveling-Wave Masers

The traveling-wave maser provides the best solution for systems which require large bandwidth and/or electronic tunability. Fig. 10 illustrates a type of traveling-wave maser circuit that utilizes nonresonant slowing for its operation. The waveguide is filled with maser material; the high dielectric constant of the material causes the wave to propagate more slowly than in an air-filled guide. The slower the wave travels, the more time is available to stimulate emission from the maser material, and thus to amplify the signal.

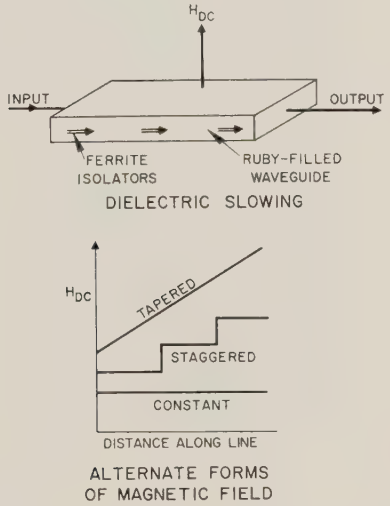


Fig. 10—Nonresonant dielectric slow-wave circuit for masers.

Ferrite isolating elements are shown embedded in the waveguide. These are required to assure that spurious reflections of the forward wave due to any cause are not amplified in the backward direction. The maser material is, in general, capable of amplifying about equally well in both directions. The presence of an amplified backward wave will lead first to variations of gain vs frequency, and ultimately to instability. Well-designed traveling-wave masers provide sufficient attenuation of backward waves so that the circuit remains stable for a short circuit at either the input or output.

The nonresonant, dielectric-loaded slowing circuit has potential for providing large bandwidths at high gains. However, unless the strength of the dc magnetic field is varied, the bandwidth is limited to a fraction of the linewidth of the paramagnetic resonance of the maser material. For ruby this linewidth is about 75 Mc. The 3-db bandwidth of a ruby traveling-wave maser that has a gain of 25 db is about 25 Mc. If the length of the circuit is increased and a staggered or tapered dc magnetic field is applied as a function of distance along the circuit, the bandwidth of the amplifier can be increased. In general, this will also require a variable pump frequency; however, at some frequencies and modes of operating the maser material, considerable bandwidth can be obtained with one pump frequency.

The nonresonant dielectric slowing circuit is not practical for use with ruby or emerald below about 10 kMc because it is too bulky. However, it shows promise for being useful with emerald and ruby in the 10- to 60-kMc band.

Fig. 11 shows two forms of traveling-wave masers which use resonant slowing. The figure at the top of Fig. 11 shows a circuit in which solid resonators of maser material are coupled together to form a band-pass filter structure. This type of circuit is capable of a large slowing factor and thus provides an efficient means for achieving large maser gain in a small structure. This is achieved, however, at the expense of the bandwidth of the structure. Thus, if one wishes to obtain large gain in a small structure, the maximum instantaneous bandwidth that can be achieved, even with staggered tuning of the maser material, is limited to the bandwidth of the structure. The electronic tuning range is similarly limited. This type of circuit should be suitable for use in the frequency band of about 3 to 30 kMc. Above this frequency range, the individual resonators may be too small to fabricate to the necessary precision; for frequencies below 3 kMc, the resonators probably will be too large for practical use.

The center figure of Fig. 11 shows a traveling-wave maser circuit which uses a comb structure. This is a resonant structure with a very large slowing factor. The

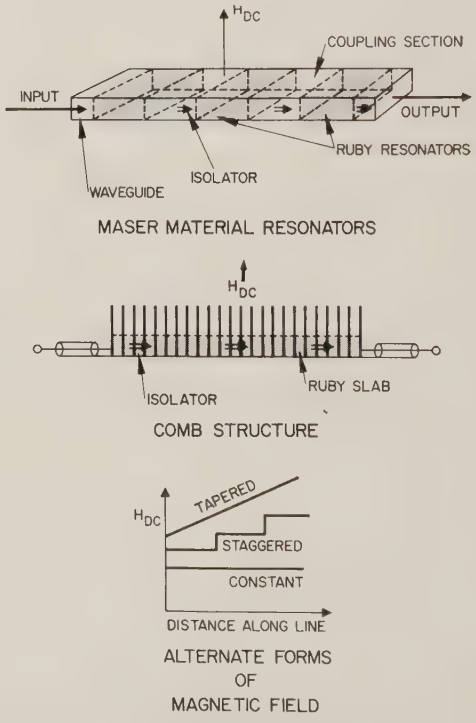


Fig. 11—Resonant slow-wave circuits for masers.

inherent loss of such structures fabricated to date has been somewhat higher than that for structures using coupled ruby resonators. This structure requires the use of a lower bath temperature (1.4°K instead of 4.2°K) to obtain satisfactory net gain per unit length. The Bell Telephone Laboratories have developed this circuit for application in the frequency range of 1 to 6 kMc.[7] A gain of 23 db and a bandwidth of 25 Mc was reported at *C* band. Electronic tuning of this circuit over the bandwidth of the structure has been achieved by varying the magnetic field and the pump frequency.

Maser Saturation and TR Protection

The presence of strong signals at the input to a maser will cause saturation with attendant reduction or complete loss of amplification. This occurs because a strong signal is capable of stimulating all of the excess maser ions stored in the high energy level to fall back to the low energy level. When the strong signal is removed, the maser recovers its gain as soon as the pump can restore the normal number of excess ions in the high energy state. For ruby this recovery time may be as long as 0.1 second. This corresponds to an interval of about 10,000 miles in the radar range. Thus, it is obvious that steps must be taken to prevent these effects in a radar.

The amount of gain reduction caused by a strong signal depends on both the amplitude and duration of the signal. Preliminary measurements indicate that typical radar-like repetitive pulse signals having amplitudes of about 100 microwatts will slightly degrade the gain of ruby masers. Similar pulses at a peak level of about 1 microwatt appear to be tolerable to the maser,[8] although they may produce other undesirable effects in the radar system.

The leak-through pulse from a conventional radar TR circuit is of the order of 1 to 10 milliwatts, and, unless special care is taken, it will be applied directly to the maser input. Thus, for any radar which uses a common antenna for transmitting and receiving, it is necessary to attenuate the TR leak-through pulse by 30 to 40 db in order to assure no degradation of maser performance.

Fig. 12 shows an early form of ferrite switch which was developed for this purpose. This switch is inserted between the TR tube and the maser used in an experimental radar. The driving coil is pulsed while the radar transmitter is on and operates with zero current during the receiving cycle. The normal mode of transmission of the waveguide is suppressed by the application of the dc magnetic field to the ferrite element. This provides more than 30 db of isolation over a 120-Mc band. The

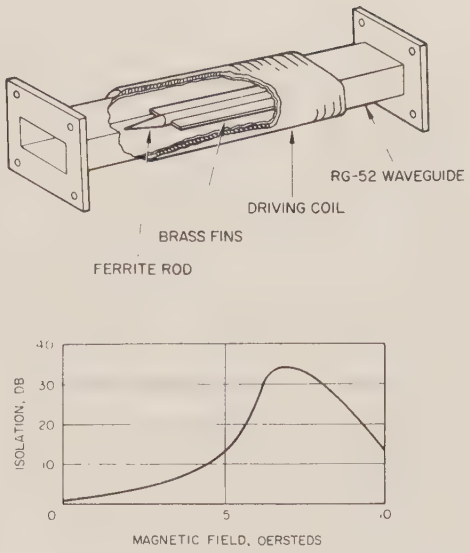


Fig. 12—Ferrite TR switch.

insertion loss during the receiving cycle is less than 0.25 db; thus, an uncooled switch contributes only about 17°K to the noise temperature of the receiver. The maser gain was not affected by the TR pulse during the experimental radar tests mentioned above. It appears reasonable to expect that improved switches, refrigerated if necessary, should contribute negligible noise to the receiver. The helium-cooled, four-port circulator shown in Fig. 9 can be switched to provide TR protection without adding appreciable noise to the receiver.

Cryogenics

Almost all of the solid-state microwave maser amplifiers operated before 1961 have used helium-nitrogen double dewars. The dewar shown in Fig. 7 is a typical example. The inner dewar is fabricated of stainless steel, one of the few practicable materials which is impervious to helium. The outer dewar, which holds the liquid nitrogen, is made of glass. The dewar of Fig. 7 is somewhat larger in diameter than many laboratory-type dewars because it was designed to contain a small permanent magnet to provide the external dc field of the maser. The maser circuitry and magnet (see Fig. 8) are immersed in the reservoir of liquid helium. The long length of the dewar is required to provide sufficient thermal impedance in the thin-walled stainless steel waveguide which connects the cooled maser elements to the input-output port, which is at normal ambient temperature.

An open-helium dewar, which is more suitable for systems use, is shown in Fig. 3. The function of the nitrogen bath of the previous dewar is performed by a radiation shield cooled approximately to liquid nitrogen temperature (77°K) by the helium gas as it escapes from the inner helium reservoir. The helium chamber is a reentrant structure which enables the device to operate in any position. The maser circuitry must be installed

[7] R. W. De Grasse, E. O. Schulz-Du Bois, and H. E. D. Scovil, "The three-level solid-state traveling-wave maser," *Bell Sys. Tech. J.*, vol. 38, pp. 305–334; March 1959.

[8] R. L. Forward, F. E. Goodwin, and J. E. Kiefer, "Application of a Solid-State Ruby Maser to an *X*-Band Radar System," Hughes Res. Labs., Malibu, Calif. Res. Rept. No. 105; June, 1959.

in the helium chamber before final assembly of the dewar. The input-output waveguide connections must be made through the small-diameter tubing which enters through the top. The waveguide must be specifically fabricated to have very high thermal impedance because of its short length; the distance between the cooled maser circuits and the input-output ports is only about 1 foot, instead of the 3-foot length of the previous dewar. A waveguide made of plastic-impregnated fiberglass with a very thin gold layer on the inner surface has been used for this purpose.

Open dewars of the types described above can be designed for holding times of up to several days. For installations in locations easily accessible to supplies of liquid helium and nitrogen, these dewars may provide satisfactory system designs. The application of masers to most operational systems, however, cannot be practically accomplished until suitable closed-helium refrigerators are available.

Other workers in the field have supplied prototype helium refrigerators for masers for use in ground electronic systems. The sizes, weights, and required power, reported for these units indicates that further engineering is needed to make them practicable for some applications. Progress of this nature is certain to occur in the near future.

A compact helium refrigerator for airborne infrared detectors and masers has been developed at the Hughes Aircraft Company. The outline dimensions for a model intended for an airborne maser are shown in Fig. 4. Fig. 13 depicts the flow diagram for this device; a photograph of an early experimental model in which the compressor unit is air cooled is shown in Fig. 14.

Helium gas is compressed to about 30 atmospheres in a two-stage compressor. This gas enters a counter-flow heat exchanger. As indicated in the flow diagram, about one third of the gas is made to enter an adiabatic expansion engine after having been cooled to 80°K. Another one third of the gas is processed similarly by the second adiabatic expansion engine. By these processes, the remainder of the gas is cooled to the critical temperature needed for Joule-Thomson cooling. After a typical cool-down time of less than 1 hour, the output of the Joule-Thomson nozzle is a two-phase mixture consisting of vapor and liquid at 4.2°K, assuming that the pressure in the return system is one atmosphere. This jet flows over a cold plate which is in thermal contact with the maser material and circuit. In this process the liquid helium is changed to a vapor at the same temperature, thus withdrawing 650 calories of heat per liter from the cold plate. The maser circuitry, which is in a vacuum on the other side of the cold plate, must be so designed and connected with the outside world that the total heat flowing into it can be absorbed by the liquid helium being vaporized on the other side of the cold plate. This arrangement requires great care in the design of the waveguides and in provision of adequate radiation shields.

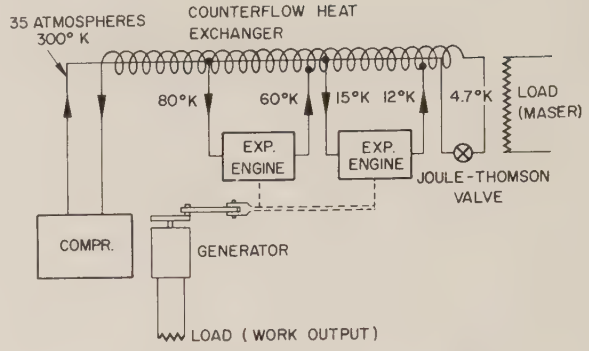


Fig. 13—Flow diagram of closed-helium refrigerator.

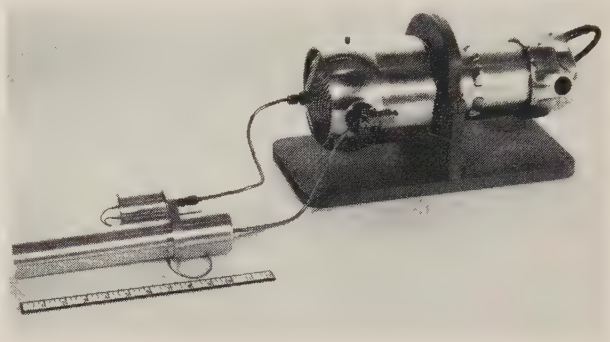


Fig. 14—Closed-helium refrigerator.

The development of an integral helium refrigerator for an airborne maser is in progress at the time of this writing. It is scheduled to be ready for flight test early in 1962.

Conclusions

In the light of the foregoing discussion, it seems reasonable to predict that the low-noise performance capabilities of masers will be of great importance in earth-to-space radar and communications systems. Ground-to-ground communications via satellites and radio and radar astronomy also present potential applications for masers. For many of these systems maser amplifiers should make possible the design of receivers for which the noise generated within the receiver is no longer a significant factor in limiting the detection range.

For the present and near future, the principal system needs and maser performance capabilities lie in the frequency range of several hundred Mc to 10 or 15 kMc. This is the useful band of ruby, which is by far the most fully developed of presently known maser materials.

Within the next few years it appears reasonable to expect that maser materials and techniques will become available for the lower millimeter-wave band. For many earthbound millimeter-wave systems, the very low noise performance of masers probably cannot be fully utilized because the external noise generated in the atmosphere is much greater in this band. However, if

millimeter-wave masers can be suitably reduced in complexity, they may play an important role because all other millimeter-wave receiving techniques now available are extremely noisy.

Appendix I

State of the Art of Ruby Masers—1961

The following is the personal estimate of the authors regarding the performance that can be achieved within the present state of the art for ruby masers operated in open cryogenic systems. In general, the numbers given are intended to indicate the order of magnitude of the various detailed performance characteristics suitable for consideration in preliminary system designs. It can be expected that many of the numbers will change significantly during the next few years. However, no attempt is made here to predict these changes.

Frequency Range	400 Mc to 15 kMc
Bandwidth	0.1 to 1.0 per cent of center frequency
Gain	20 to 40 db
Tuning Range	10 to 20 per cent
Noise Temperature	10 to 30°K
Dynamic Range	100 db
Onset of Saturation (low duty cycle pulses)	10^{-5} watts peak
Recovery Time	0.1 second
Weight of Maser-Dewar Unit (excluding pump-power supply)	
1) with permanent magnet in dewar	15 to 50 lbs.
2) with external permanent magnet	100 to 500 lbs.
Weight of Pump-Power Supply	25 to 100 lbs.
Open Helium Dewars	
1) holding time	8 to 35 hours
2) size	0.3 to 2.5 cu. ft.
3) weight	10 to 45 lbs.
4) helium boil-off rate	0.1 to 0.3 liters per hour
5) helium capacity	2 to 10 liters

Appendix II

Tentative Performance for Closed-Cycle Helium Refrigerators

The following are intended as order-of-magnitude predictions of the performance of helium refrigerators in 1962–63 suitable for preliminary planning purposes for future electronic systems which may use masers. They represent the authors' judgment based on rather complete information on cryogenic developments at the Hughes Aircraft Company and limited information on similar work elsewhere.

Refrigerators for Ground Installations	
Cooling Capacity at 4.2°K	$\frac{1}{4}$ to 1 watt
Size	3 to 10 cu. ft.
Weight	100 to 500 lbs.
Power Required	1 to 5 kw
Life (between major overhauls)	
1) cryostat	1000 hours
2) compressor	5000 to 10,000 hours
Refrigerators for Air- or Missile-Borne Applications	
Cooling Capacity at 4.2°K	$\frac{1}{4}$ to $\frac{1}{2}$ watt
Size	1 cu. ft.
Weight	35 to 50 lbs.
Power Required	1 to 2 kw
Life (between major overhauls)	1000 hours

Acknowledgment

Numerous members of the Technical Staff of the Hughes Aircraft Company have contributed to the work described in this paper. Special credit must be given to Dr. T. H. Maiman and his group, who conducted the basic research on solid-state maser amplifiers under Signal Corps and Hughes sponsorship, and to Dr. H. Barhydt and his group for their work on cryogenics.

The Electron Beam Parametric Amplifier as a Radar System Component*

R. ADLER†, FELLOW, IRE, AND W. S. VAN SLYCK†, SENIOR MEMBER, IRE

Summary—**Practical experience is now available with the Electron Beam Parametric Amplifier (EBPA) used as a low-noise preamplifier in radar systems. Principles of operation of the EBPA are briefly recapitulated. Among its desirable characteristics are unilateral behavior, high stable gain and a double-channel noise figure (NF) of 1 db or less; its bandwidth, input and output impedance remain constant as gain is adjusted. An *L*-band amplifier is described in detail. A discussion of various ways of handling the idler channel follows. The technique of in-pass-band pumping achieves an effective NF of 2.5 db or better in radars with conventional detection. Complications arise with Moving Target Indicator (MTI) systems; remedies are described.**

The paper reports on field tests in which these techniques were used, including a few flight tests. The results show substantial improvements in sensitivity and range. They also spotlight the ability of the EBPA to tolerate high overload and to act somewhat like a TR switch.

Introduction

A LITTLE more than two years have passed since the first report on successful operation of a low-noise Electron Beam Parametric Amplifier (EBPA) appeared in the technical literature.[1] Since then, several papers have been published which dealt with the theory of the amplifier,[2] compared its performance with that of diode amplifiers[3] and discussed further developments.[4] Little has been said about applications.

A good deal of practical experience is now available regarding the performance of the EBPA as a low-noise preamplifier in radar systems. This paper reports on that experience. Because the EBPA is not yet widely known, a brief review of its operation may be in order.

Fig. 1 illustrates, in highly schematic form, the essential functions. The amplifier consists of three separate sections: the input coupler, the quadrupole, and the output coupler. An electron beam traverses these three sections. It is held together by a homogeneous magnetic field.

The input coupler is made up of a pair of parallel plates, arranged somewhat like the deflection plates in a cathode-ray tube. In response to a signal, it generates transverse electron motion. The intensity of the magnetic field is adjusted to produce cyclotron resonance for electrons at the center frequency of the signal band. An incoming signal then causes electrons to spiral with increasing radius as they travel through the coupler region; eventually the entire energy of the signal is converted into energy of electron motion.

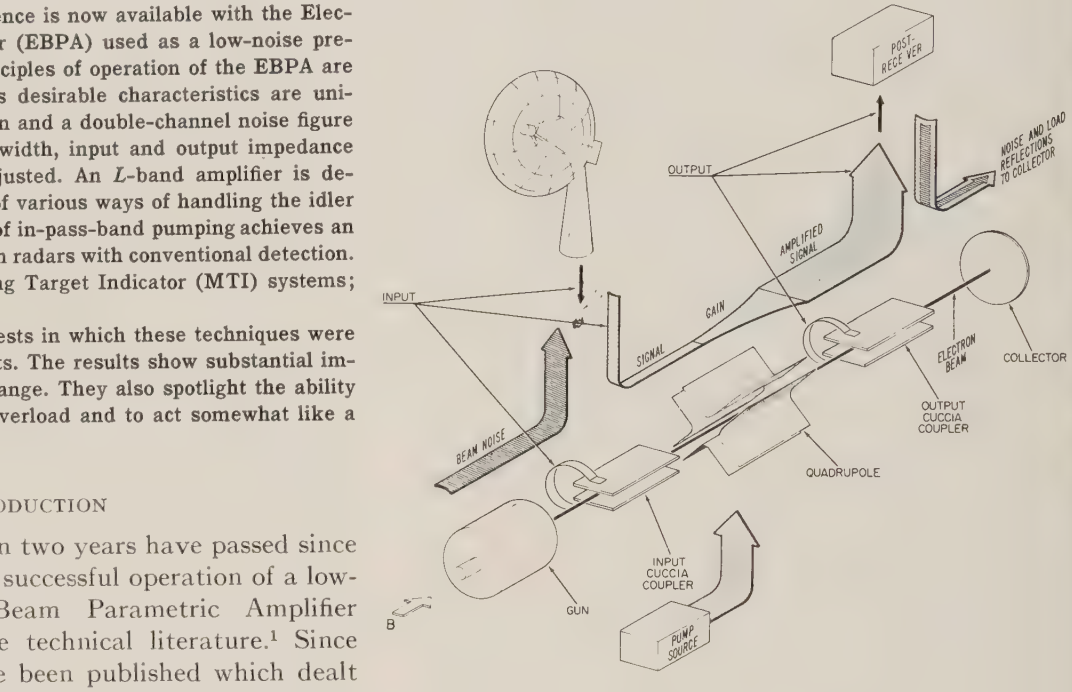


Fig. 1—Noise and signal flow in the EBPA.

In the output coupler, the inverse process takes place; the transversely-moving beam induces current in the coupler plates, and if they are correctly loaded, all the signal power carried by the beam is extracted. A tube containing only input and output couplers thus acts like an isolator.

The mechanism which permits a load properly matched to the output coupler to extract a signal previously impressed upon the beam also operates in the input coupler; there it removes noise from the beam and directs it toward the signal source. The combined effect of the two couplers is equivalent to that of two three-port circulators, connected by a unilateral energy transfer device. This equivalence, fundamental to the behavior of the EBPA, is illustrated in Fig. 2.

The quadrupole is inserted between input and output couplers. A parametric process causes the radius of the spiraling electron motion to increase exponentially. This is accomplished by driving the quadrupole with a

* Received by the PGMIL, January 17, 1961.
† Zenith Radio Corp., Chicago, Ill.
[1] R. Adler and G. Hrbek, "A low noise electron beam parametric amplifier," PROC. IRE, vol. 46, pp. 1756–1757; October, 1958.
[2] R. Adler, G. Hrbek, and G. Wade, "The quadrupole amplifier, a low noise parametric device," PROC. IRE, vol. 47, pp. 1713–1723; October, 1959.
[3] C. B. Crumly and R. Adler, "Electron beam parametric amplifiers," *Electronic Ind.*, vol. 18, pp. 73–76; November, 1959.
[4] T. J. Bridges and A. Ashkin, "A microwave Adler tube," PROC. IRE, vol. 48, pp. 361–363; March, 1960.

pump signal of approximately twice the cyclotron frequency. Because the signal and cyclotron frequencies are approximately equal, the ratio between pump frequency and signal frequency is also about 2:1. The action of the quadrupole has been explained in detail in the literature.[2]

Many of the characteristic properties of the EBPA are easily understood. Most important perhaps is its unilateral behavior; this results from the one-way action of the electron beam which, like a conveyor belt, transfers the signal from input coupler to amplifying section to output coupler. There is no return path, no chance of feedback or instability.

The circulator-like behavior of input and output couplers is important in practice; it renders the impedances of these couplers independent of each other, and they do not change when the pump is turned on or off. The coupler circuits alone determine the bandwidth of such an amplifier; quadrupole construction and pump power determine its gain. Bandwidth and gain are thus independent of each other.

The noise figure (NF) of an EBPA, measured with a broad-band noise source, should in principle be unity, or 0 db.[5] In practice, unavoidable circuit losses plus small residual contributions from the electron beam add up to 1 db or less. Figures as low as 0.6 db have been measured.

L-Band Amplifiers

The radar applications covered by this report all employ frequencies in the vicinity of 1300 Mc. Fig. 3 shows the RF head of an amplifier for this frequency range.[6] Most of the box is taken up by a solenoid; inside the solenoid, the EBPA tube is mounted in a capsule which provides a broad-band match between the clearly visible input and output lines and the corresponding couplers hidden inside the tube.

Fig. 4 illustrates the signal transfer mechanism; the Cuccia coupler plates, show on the extreme right with the deflected electron beam sketched in, are tuned to the center of the signal band by a single-turn inductance mounted inside the tube envelope close to the glass. Outside the envelope there is a second tuned circuit, magnetically coupled to the single turn inside. The external-tuned circuit is matched to 50 ohms by means of a low-impedance quarter-wave line section.

Fig. 5 shows a 1300-Mc tube, the two external coupling devices and the metal shell which surrounds them. Mutual coupling between internal and external circuits is adjusted by rotating the two external coupling devices.

The external-tuned circuit, the internal-tuned circuit,

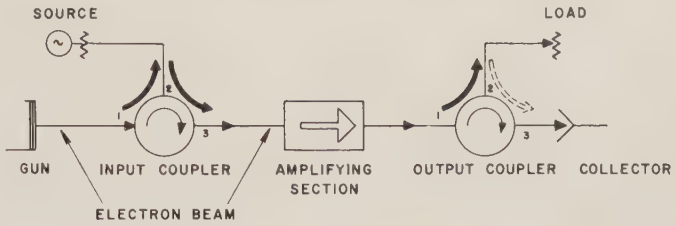


Fig. 2—Circulator action of the input and output couplers.

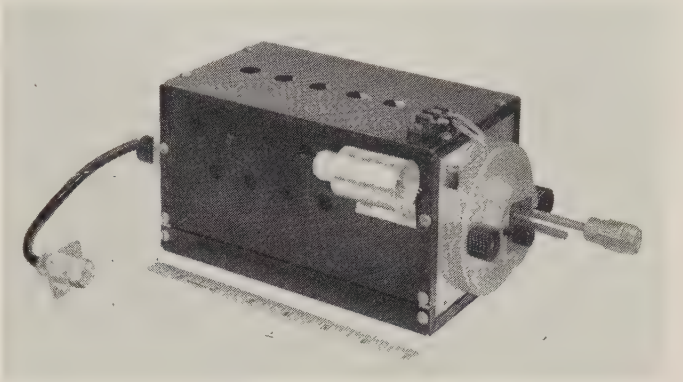


Fig. 3—*L*-band amplifier— RF unit.

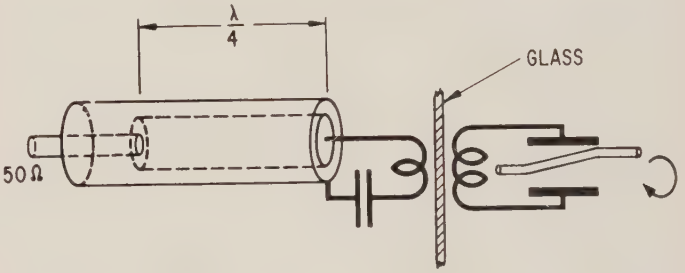


Fig. 4—Inductive transfer of signal energy through the tube envelope.

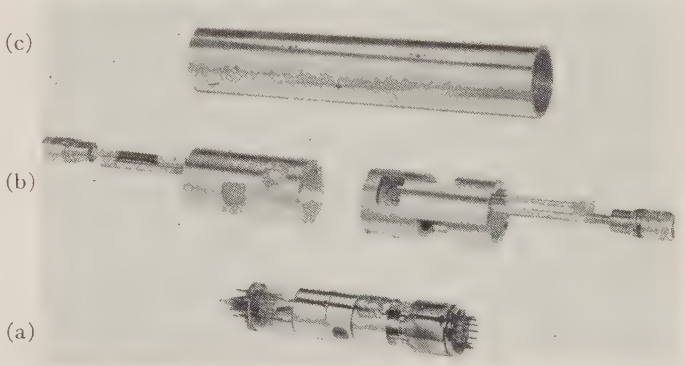


Fig. 5—*L*-band capsule. (a) EBPA tube. (b) External coupling devices. (c) Shell for mounting in solenoid.

and the resonant electron beam associated with each coupler form simple band-pass filters which are adjusted to provide a good impedance match between the electron beam and a 50-ohm line over a band of about 80 Mc. The NF is best in the center of this band (1 db, often better) and remains good (1.0–1.5 db) over about 50 Mc, rising near the edges of the 80-Mc band.

[5] R. Adler, "Parametric Amplification of the Fast Electron Wave," presented at the Conf. on Electron Tube Res., Berkeley, Calif.; June, 1957. See also Proc. IRE, vol. 46, pp. 1300–1301; June, 1958.

[6] C. B. Crumly and R. L. Cohoon, "A Low Noise Electron Beam Parametric Amplifier at *L*-Band," presented at the Electron Devices Conf., Washington, D. C.; October, 1960.

Handling of the Idler Channel

Any report on the use of an EBPA in a radar receiver must be preceded by the answer to a most important question: What was done about the idler? The EBPA is a parametric device. Like all other parametric devices, it amplifies signals not only at the nominal signal frequency but also at the idler frequency, the difference between pump and signal frequencies. With an EBPA, the pump frequency is never far from twice the signal frequency; in fact, the idler frequency must lie within the pass band of the input coupler to ensure removal of idler frequency beam noise. Signal and idler may be as much as 50 Mc apart, but this is still only a few per cent at 1300 Mc. Fig. 6 gives an example for the positions of signal, idler, and one-half pump frequency within the input coupler pass band.

Signals and noise can enter the amplifier equally well on signal and idler channels. Signal energy may be present in only one of these channels; noise unavoidably exists in both. Even if the amplifier were perfect and contributed no noise whatever, the thermal noise from the idler termination would still appear in the output; this noise must be counted against the EBPA, since no similar noise source exists in conventional amplifiers. This extra noise does not show up when the NF is measured with a broad-band noise source, because such a source feeds power into both channels [Fig. 7(a)]. The case of signal energy in one channel only, with the idler channel terminated separately in a resistance R_0, is illustrated in Fig. 7(b); if R_0 is at room temperature, 3 db must be added to the broad-band (or double-channel) NF of the EBPA to obtain its effective single-channel NF.

Sometimes an antenna is aimed at a point in the sky high above the horizon. In such cases it is best to use no filter at all and let the sky act as idler termination [Fig. 7(c)]. If the EBPA can be located close to the antenna, and if the increase in idler termination temperature produced by duplexer and TR losses can be kept low, single-channel performance is very good. Such an arrangement is particularly interesting because the EBPA needs little TR protection; more about this later.

In many radars, on the other hand, the antenna is aimed at the horizon, and often duplexer and TR losses are large enough to bring the effective noise temperature of the idler termination close to 290°K. The arrangement of Fig. 7(c) is then not particularly good. Is there another solution?

If the noise contribution from the idler channel could somehow be completely suppressed, the amplifier would perform in accordance with its double-channel NF of 1 db or so. When an idler channel brings in noise but no signal power, the signal-to-noise ratio is cut in half. What happens when the pump frequency is so chosen that signal and idler channels coincide, so that both carry signal power? Because the phase of a radar echo is not known in advance, it is not possible to arrange for the signal contributions from the two channels to cooperate under all circumstances; instead, they will combine in random phase relationship. Still, on the average, the signal recovered by this method is larger[7] than if the idler channel carried no signal power; the required averaging process occurs on the scope face or in the eye whenever several echos from a given target arrive in quick succession. The technique (Fig. 8) has become known as in-pass-band pumping; the pump frequency need not be accurately twice the transmitter frequency,

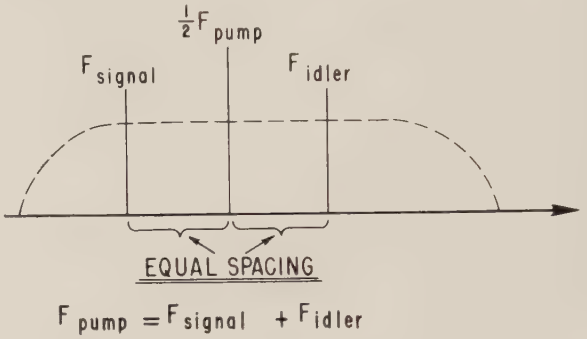


Fig. 6—Suitable positions of significant frequencies in the EBPA pass band.

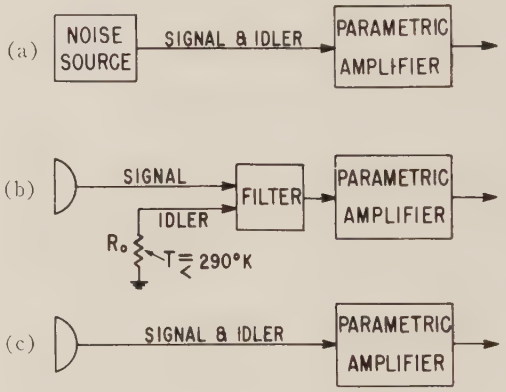


Fig. 7—Different ways of handling the idler channel: (a) Broad-band noise measurement. (b) Separate idler termination. (c) Idler terminated in sky.

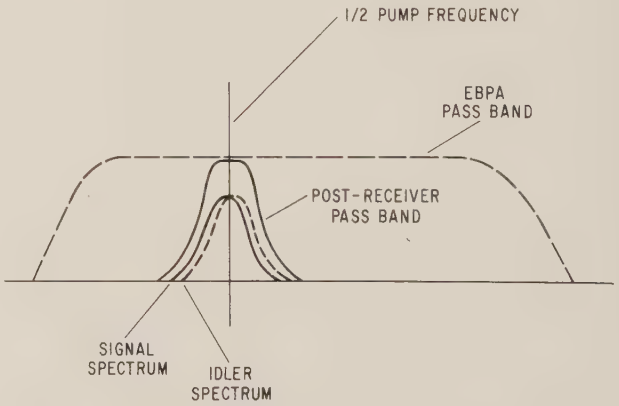


Fig. 8—In-pass-band pumping.

[7] R. Adler, "Electron beam parametric amplifiers with synchronous pumping," *Proc. Symp. on the Application of Low Noise Receivers to Radar and Allied Equipment*, Lincoln Lab., M.I.T., Lexington, Mass., vol. 3, pp. 177–197; November, 1960.

even though frequency doubling has proved a convenient method for generating the pump signal.

It has been found that in-pass-band pumping produces an effective single-channel NF only 1.0 to 1.5 db poorer than the double-channel NF of the EBPA. Using tubes available today with a double-channel NF of 1 db results in single-channel performance corresponding to 2.0 to 2.5 db. The tubes employed in the experiments reported here were built earlier and were about 0.2 to 0.5 db poorer.

Field Test Results

Table I lists experimental results obtained in five different *L*-band radar installations and on one TACAN site. Conventional detection and in-pass-band pumping were used throughout.

TABLE I

Equipment	Noise Figure Improvement (double-channel over original)	MDS Improvement	Source Termination
FPS-8 Radar Rome, N. Y.	7.2	9	antenna
ARSR-1 Radar Wayland, Mass.	4.9	4.0	dummy load
MPS-11 Radar Glenview, Ill.	8.1	7.0	antenna
AASR-1 Radar North Bay, Ont., Canada	5.5	4.5	dummy load
GRN-9B TACAN Patuxent River, Md.	6.3	4.3	dummy load
MPS-11 Radar Jacksonville, N. C.	8.3	7–10	antenna

The improvement in MDS (minimum discernible signal) was determined on an *A* scope by operators on the various radar sites. Specialized test equipment was used on the TACAN. In the first and last case listed, the observed MDS improvement was substantially larger than might have been expected from the NF improvement; this difference was probably caused by low antenna temperature.

In the location listed last on the chart, controlled flight tests were made. Typically, a test plane few along a fixed radial at constant altitude. Each time the antenna swept across the selected azimuth, the spot painted on the PPI scope was graded by the operators as follows: 1) persistent for the entire scan; 2) for only half the scan; 3) barely visible. Grade and distance were recorded. The results were quite reproducible and there was no noticeable fluctuation from scan to scan. In these tests, a 70 per cent increase in range after insertion of the EBPA was observed on one day, and a 45 per cent increase on another. The 70 per cent figure corresponded to in-pass-band pumping; the 45 per cent figure was obtained with the pump detuned sufficiently to separate the idler channel completely from the signal channel. Similar comparisons in other cases, but using *A*-scope presentation and simulated signals, gave a difference of 2.5–3.0 db between in-pass-band and out-of-pass-band pumping. While these observations would indicate that nearly all the detrimental effect of the idler is removed by in-pass-band pumping, a more conservative estimate assumes that a residual impairment exists. This may be 1.0 to 1.5 db, as mentioned before.

In the experiments described, in-pass-band pumping was not used in connection with MTI circuits. The presence of an extra signal from the idler channel, bearing an arbitrary phase relation to the main signal, would upset MTI circuits. Several of the radar installations where experiments were made contained a time-gated mode switch. This is usually operated in such a manner that the MTI circuits are activated up to a certain maximum range; then a switch-over is made to conventional detection in order to take advantage of the better sensitivity of this mode. An adjustable timing circuit triggers the change-over a given number of microseconds after each transmitter pulse.

A practical arrangement which has proved highly successful in the field tests takes advantage of this range controlled switching (Fig. 9). The change-over trigger is employed to perform a second, simultaneous switching operation. At close range, while the receiver is in MTI mode, the pump frequency is set a few megacycles away from synchronism, thus placing the idler outside the IF band. When the receiver is switched to conventional detection at longer range, the pump frequency is shifted to near-synchronism. With an EBPA having a double-channel NF of 1 db, the effective single-channel NF is then 4 db or better at close range and 2.5 db or better at long range. This system maintains full subclutter visibility during MTI operation while providing optimum visibility for weak signals at long range.

A variant of this arrangement takes advantage of an unusual property of the EBPA: With the pump turned off, the EBPA becomes a simple isolator with about 1-db insertion loss. The gain of the EBPA can thus be varied smoothly from -1 db to the desired maximum of 20 or 30 db by gradually increasing the pump power; this has no effect on input and output impedances. A timing voltage, again derived from the transmitter pulse, is employed to turn the pump on gradually, starting at some intermediate range where NF improvement becomes important while fixed targets become weaker. This mode of operation automatically provides a certain amount of sensitivity-time control.

In some cases, it is sufficient to turn the pump on and off abruptly, instead of gradually. This simple arrangement is shown in Fig. 10.

Can an EBPA be used in connection with a coherent MTI radar in such a way that signal and idler channels coincide and cooperate? This method, often referred to as synchronous pumping, was suggested several years ago by K. G. Eakin, then of Rome Air Development Center. The idea was to synchronize the pump generator

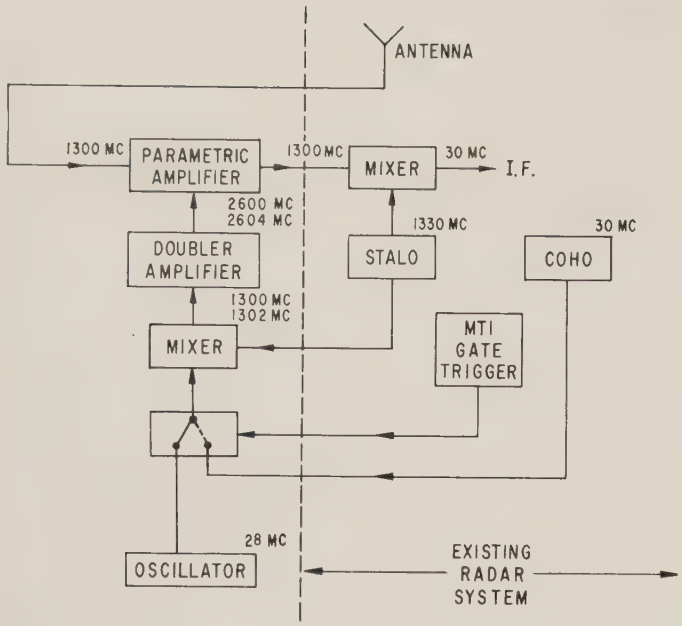


Fig. 9—Mode switching system for MTI radar.

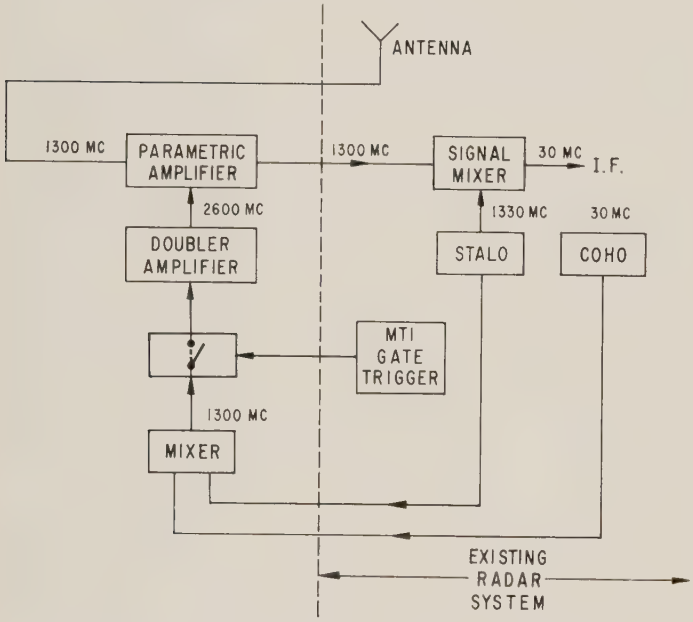


Fig. 10—Simplified mode switching for MTI radar.

with the second harmonic of the transmitted signal frequency and to take advantage of a properly which a synchronously pumped parametric amplifier shares with the phase detector used in MTI circuits: both respond to one signal component (say, cosine θ), but reject the corresponding quadrature component (sine θ). Eakin showed that when the pump generator is properly phased, the entire output from the parametric amplifier, including signal and idler components, is accepted by the phase detector; the sine θ component rejected by the parametric amplifier would be suppressed by the phase detector in any event. For weak signals (below limiting level) one would therefore expect normal MTI operation with a sensitivity in accordance with the double-channel NF of the parametric amplifier, without impairment.

Such a system was tested in Rome, N. Y., early in 1960.[8] Eakin's prediction with respect to performance under weak signal conditions was fully confirmed. Later, however, it was found that a substantial impairment of subclutter visibility occurs whenever fixed target clutter exceeds limiting level. Since MTI circuits do most of their useful work at close range where fixed target echos are quite intense, the system has little practical utility when used with a limiting IF amplifier. It has been suggested that a substantial improvement might result from the use of a logarithmic IF; this has not yet been tested.

Up to this point, attention has been focused on MDS improvement and on pumping methods. However, there are other experimental observations which deserve to be recorded; they bear on the question of whether an EBPA should be used in a specific installation.

The amplifiers have shown their expected bandwidth of about 50 Mc with good NF. Gain of at least 30 db has always been available. Stability has been very good; amplifiers installed on radar sets have been turned on and off many times over periods of weeks without noticeable change in performance. While the total number of hours accumulated in any of the test installations was not large enough to qualify as a life test, amplifiers have operated in the laboratory for more than 5000 hours without observable change.

The EBPA has a number of unusual properties which may become important in some instances. For example, phase stability was found to be better than $\pm 2°$ over periods of many hours. Gain stability is primarily a function of pump power regulation; at 20-db gain there is a variation of 1.3 db for 10 per cent change in pump power. Two amplifiers operated from a common power supply and common pump remained within ± 0.5 db of each other, even though pump power was varied to produce a ± 5-db gain adjustment. These properties are desirable in phased arrays. The high stability is maintained under environmental extremes.

The input-vs-output characteristic of the EBPA is unusually linear up to the point at which the excursion of the beam causes it to touch the output coupler. Thus linearity continues up to saturation. This is quite unusual for a low-level amplifier.

Once saturation is reached (about -10 dbm in 1300-Mc tubes), the output from the EBPA decreases again. The crystal mixer which follows can never be damaged, regardless of signal level. Whether the EBPA is turned on or turned off, this protection remains. It can result in a substantial reduction in down-time.

The protective feature just mentioned is particularly useful because the input structure of an EBPA is quite

[8] "High-frequency radar amplifier," PROC. IRE (News), vol. 48, pp. 174A–178A; June, 1960.

rugged and can dissipate substantial transmitter leakage. A large signal causes the beam to spread and strike the input coupler plates. Since beam current is only about 50 μa, this is harmless. Once the beam disappears from its assigned location inside the input coupler, the input impedance becomes highly reactive; the standing-wave ratio of the input coupler changes from unity at low input levels to about 10 or more at overload levels. Thus, the EBPA not only withstands large amounts of incident power, but also reflects most of this power. Yet, after the overload is removed, it returns instantly to normal. These are the essential characteristics of a TR device. A substantial part of the TR function thus can be absorbed by the EBPA, with consequent improvement in system NF as well as reduced maintenance, down-time, and cost.

In this connection, it may be of interest to relate an incident which occurred on an MPS-11 radar at Glenview, Ill. This installation was operated for approximately one hour with the EBPA in place, with a broken TR tube. Operation was not normal at close range, but appeared almost normal at long range. After the trouble was located, the TR tube was replaced and everything went back to normal; there was no change in the NF of the EBPA nor of the crystal mixer behind it.

New Developments

The field tests reported in this paper were all made in ground installations. Power consumption, weight and size of the EBPA are of no particular interest in these installations. For other uses, weight and size are undesirably large. Furthermore, while the present EBPA tubes are quite rugged, they are not rugged enough for many applications. Their main disadvantage, perhaps, is that they are not tunable; each tube is equipped with fixed-tuned internal circuits which determine the signal frequency band.

A metal-ceramic version of the EBPA (Fig. 11) is now in the final stage of its development. This is a much more compact structure which does not include built-in inductances. Instead, low-inductance leads to the couplers and quadrupole are brought out radially through the ceramic envelope, permitting a great variety of external circuits. These tubes may, for instance, be tuned over a 200-Mc band centered at 1300 Mc by a simple tuning mechanism. They can be build alike for a wide range of frequency bands. Naturally, the metal-ceramic structure is extremely rugged with respect to shock, vibration and temperature. Because there is no need for two tuned circuits to be associated with each coupler as is the case with the glass tube, the circuit losses in the input coupler are somewhat less in the ceramic tube, resulting in a further improvement in NF.

In considering the possibility of EBPA tubes for *S*-band and higher frequencies, it should be remembered that an amplifier operating at 4140 Mc was described some time ago.[4] Tubes of this kind can probably be built up to the *X*-band region.

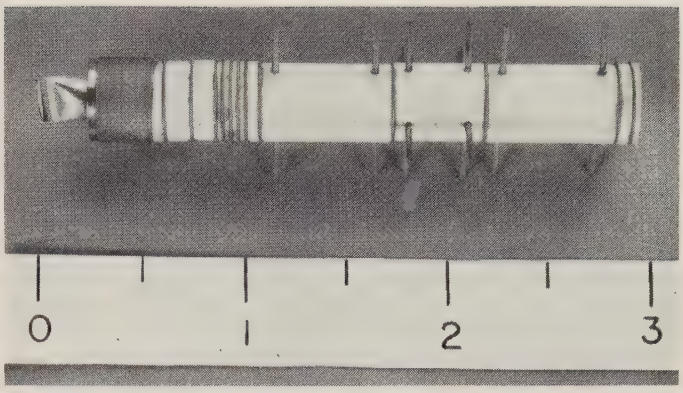


Fig. 11—Metal-ceramic EBPA tube.

It would be highly desirable to dispense with the idler by making the pump frequency many times higher than the signal frequency, an arrangement commonly used in diode amplifiers. Removal of the idler would be useful from the standpoint of use in MTI circuits, even though the advantage obtained with conventional detection would be only slight.

The task of constructing a low-noise nondegenerate EBPA is not easy. Early experimental results along this line were reported last year;[9] the tube was pumped with 2000 Mc and amplified a signal frequency of 400 Mc. Its noise performance was not good. Experimental and theoretical work has continued since; at least one of the sources of the unexpected additional noise was found and eliminated. This, however, is not the whole story; a substantial effort is now underway to determine the nature of the remaining noise and the possibility of its removal. It is too early to say whether the nondegenerate EBPA will eventually match the low-noise performance of its present near-degenerate prototype.

Conclusion

Summing up, the EBPA has proved its worth as a radar preamplifier; its high stable gain, its ability to cope with heavy overload, to protect the mixer crystal, and to recover instantly have proved valuable. Difficulties which stem from the presence of the idler channel have been largely overcome by the technique of in-pass-band pumping. Field tests show that for conventional detection and PPI presentation, the effective noise figure achieved with this simple technique is 2.5 db or better. Complications arise in connection with MTI circuits; they can be circumvented if a somewhat poorer noise figure during MTI operation is acceptable.

The EBPA is probably the only low-level amplifier which doubles as a TR device.

A new metal-ceramic tube, designed for operation at *L* band and below, promises increased ruggedness. It also gets away from built-in tuned circuits and permits the construction of tunable amplifiers.

[9] C. B. Crumly, "Recent Advances in Nondegenerate Cyclotron Wave Amplifiers," presented at the Conf. on Electron Tube Research, Seattle, Wash.; June, 1960.

Design Considerations for Parametric Amplifier Low-Noise Performance*

C. R. BOYD, JR.†, MEMBER, IRE

Summary—The basic characteristics of parametric amplifiers are reviewed briefly, with particular emphasis on the limitations on low-noise performance resulting from diode losses. Normalized curves showing the minimum excess temperature of a parametric amplifier for various spectrum arrangements are presented, where the normalizing factor is the degenerate-mode gain cutoff frequency. Experimental evidence consisting of noise figure measurements on an *X*-band amplifier is used to relate this cutoff frequency to an easily and commonly measured diode quality factor. Design considerations for a hypothetical *L*-band parametric amplifier are discussed to illustrate the implications of the noise analysis.

I. Introduction

An impressive amount of literature[1,2] has been published on the subject of parametric amplifiers in the past few years. Most of this literature has been either concerned with or stimulated by the possibilities of low-noise performance from such circuits. However, a quantitative formulation of the performance limitations of semiconductor-diode amplifiers has been developed only comparatively recently.[3–5]

In this paper, the basic characteristics of parametric amplifiers are reviewed briefly, and the limitations on low-noise performance resulting from diode losses are examined. Normalized curves showing the minimum excess temperature of a parametric amplifier for various spectrum arrangements are presented, where the normalizing factor is the degenerate-mode gain cutoff frequency. Experimental evidence consisting of noise figure measurements on an *X*-band amplifier is offered as a means of relating this frequency to the usual "cutoff frequency" as measured on an unpumped diode.[6] Finally, design considerations for a hypothetical *L*-band parametric amplifier will be discussed to illustrate the implications of the noise analysis.

II. Basic Considerations for Lossy Parametric Amplifiers

When used in parametric amplifiers, the semiconductor diode is customarily represented as a voltage-variable capacitor in series with a fixed resistor. Measurements of contemporary diodes in the useful reverse-biased region have shown experimentally that the assumption of a constant series resistance, R_s, is reasonably valid,[7] and this is the principal justification for assuming such an equivalent circuit. In the parametric amplifier, the diode is "pumped" by a frequency, f_p, and excited by a signal frequency, f_0. In addition, the circuit loading on the diode at the difference frequency, $f_{-1}=f_p-f_0$, is carefully controlled.

Since the amplitude of the pumping voltage is generally very large compared to the information-carrying frequencies, the capacity $C(v)$ of the diode is assumed to vary only as a result of the former. Thus the pumping circuit exhibits essentially nonlinear, large-signal behavior, so that variations at harmonic rates are produced. If the capacity-voltage characteristic and the time variation of voltage are known, it is possible to generate the waveform of capacity vs time. This can be done analytically[8] if the capacity-voltage law is known, or graphically[9] if only an experimental curve is available. With the use of a single-frequency pumping voltage, the variation of capacity in time will be periodic at the pumping rate, and is expressible as a Fourier series involving harmonics of the pumping frequency.

For the sake of convenience in handling the diode losses in subsequent analysis, it is desirable to use elastance as the variable parameter rather than capacitance.[5] The variation in time of the diode elastance $\sigma(v)=1/C(v)$ may then be represented by the complex Fourier series,

$$\sigma(t) = \sum_{m=-\infty}^{\infty} \sigma_m e^{jm(\omega_p t+\phi_p)}, \tag{1}$$

where $\omega_p=2\pi f_p$ and ϕ_p is a reference phase angle. If the elastance is assumed to be hysteresis-free, the $\sigma(t)$ waveform will have even symmetry. The reference phase angle ϕ_p is then chosen such that the σ_m coefficients are pure real, with $\sigma_{-m}=\sigma_{+m}$ a necessary condition. By inference, then, the instantaneous applied pumping voltage $v_p(t)$ has been taken as

$$v_p(t) = V_p \cos(\omega_p t + \phi_p), \tag{2}$$

where V_p is its peak value.

* Received by the PGMIL, January 17, 1961.
† Electronics Lab., General Electric Co., Syracuse, N. Y.

[1] W. W. Mumford, "Some notes on the history of parametric amplifiers," PROC. IRE, vol. 48, pp. 848–853; May, 1960.

[2] E. Mount and B. Begg, "Parametric devices and masers: an annotated bibliography," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8, pp. 222–243; March, 1960.

[3] J. C. Greene and E. W. Sard, "Optimum noise and gain-bandwidth performance for a practical one-port parametric amplifier," PROC. IRE, vol. 48, pp. 1583–1590; September, 1960.

[4] K. L. Kotzebue, "Optimum noise performance of parametric amplifiers," PROC. IRE, vol. 48, pp. 1324–1325; July, 1960.

[5] P. Penfield, Jr., "The High-Frequency Limit of Varactor Amplifiers," Signal Corps Contract DA-36-039 Sc-78236, Third Quarterly Progress Rept.; May 1 to July 31, 1959.

[6] N. Houlding, "Measurement of varactor quality," *Microwave J.*, vol. 3, pp. 40–45; January, 1960.

[7] A. Uhlir, Jr., "The potential of semiconductor diodes in high frequency communications," PROC. IRE, vol. 46, pp. 1099–1116; June, 1958.

[8] M. Uenohara, "Noise consideration of the variable capacitance parametric amplifier," PROC. IRE, vol. 48, pp. 169–179; February, 1960.

[9] S. Sensiper and R. D. Weglein, "Capacitance and charge coefficients for parametric diode devices, "PROC. IRE, vol. 48, pp. 1482–1483; August, 1960.

The voltage drop across $\sigma(t)$ which results from complex currents I_0 and I_{-1} flowing at frequencies ω_0 and ω_{-1} is given by

$$v(t) = \sigma(t) \int i(t)dt = \mathrm{Re}\left[\left(\frac{I_0}{j\omega_0} e^{j\omega_0 t} + \frac{I_{-1}}{j\omega_{-1}} e^{j\omega_{-1} t}\right) \cdot \sum_{m=-\infty}^{\infty} \sigma_m e^{jm(\omega_p t + \phi_p)}\right]. \quad (3)$$

The voltage components of the above relationship which have frequencies other than ω_0 and ω_{-1} are important only if they produce current flow at those frequencies. If the diode is open-circuited at frequencies other than ω_0 and ω_{-1}, the existence of these other small-signal voltage components is merely incidental, and they may be dropped from the analysis. What is left is the following:

$$v(t) = \mathrm{Re}\left[\left(\frac{\sigma_0}{j\omega_0} I_0 - \frac{\sigma_1}{j\omega_{-1}} I_{-1}{}^* e^{j\phi_p}\right) e^{j\omega_0 t} + \left(\frac{\sigma_1}{j\omega_0} I_0 e^{-j\phi_p} - \frac{\sigma_0}{j\omega_{-1}} I_{-1}{}^*\right) e^{-j\omega_{-1} t}\right]. \quad (4)$$

The bracketed terms in the above relationship may be recognized as complex voltages, and defined as V_0 and $V_{-1}{}^*$, at ω_0 and ω_{-1}, respectively. Once this has been done, it is possible to formalize the V-I properties of the circuit in an impedance matrix:

$$\begin{bmatrix} V_{-1}{}^* \\ V_0 \end{bmatrix} = \begin{bmatrix} j\frac{\sigma_0}{\omega_{-1}} & -j\frac{\sigma_1}{\omega_0} e^{-j\phi} \\ j\frac{\sigma_1}{\omega_{-1}} e^{j\phi} & -j\frac{\sigma_0}{\omega_0} \end{bmatrix} \begin{bmatrix} I_{-1}{}^* \\ I_0 \end{bmatrix}. \quad (5)$$

The main-diagonal terms of this matrix represent the equivalent static series reactance of the mixer, which appears at both the signal and upper sideband ports. The effect of pump level on σ_0 and, hence, on this reactance is generally of secondary importance, so that the over-all effect is very nearly that of an ordinary capacitive reactance. However, the off-diagonal terms, which describe the coupling between ω_0 and ω_{-1}, are strongly dependent on the pumping level because of the σ_1 term. The nonreciprocal nature of the interaction is demonstrated by the amplitude inequality of the off-diagonal terms, which is simply an expression of the Manley-Rowe relation:[10]

$$\frac{W_{-1}}{W_0} = \frac{f_{-1}}{f_0}. \quad (6)$$

For the single-diode system under consideration here, the assumption of incommensurate signal and pump frequencies permits an arbitrary choice of ϕ_p; hence, it will be conveniently set equal to zero.

The impedance matrix of (5) is representative thus far of the lossless case only, but it may be converted into the lossy case matrix by the addition of R_s terms into the main-diagonal elements. However, a more useful representation for the parametric amplifier interaction than the impedance matrix is the scattering matrix. This conclusion follows from the definition of amplifier input and output as incident and reflected waves, and the amplifier gain as the power reflection coefficient associated with a transmission line terminated in a pumped diode. The general rules for formulating the scattering matrix of a network whose impedance matrix is known have been summarized by Dicke.[11] In a recent paper,[12] the author has outlined in detail the procedure for computing the scattering matrix of a general parametric mixer and for subsequently evaluating the gain and noise figure performance of such a mixer through the use of its scattering matrix elements.

Briefly, this technique involves first the derivation of the lossless mixer scattering matrix S from its properly normalized impedance matrix, using the matrix relationship:

$$S = (Z - 1)(Z + 1)^{-1}. \quad (7)$$

In this initial derivation, the bulk resistance, R_s, is lumped in with the signal and sideband loads. The R_s terms are then separated from the external loads, and auxiliary scattering matrices are written to describe the scattering between the R_s, external load, and lossless mixer portions of the circuit, at both the signal and sideband frequencies. The elements of the auxiliary matrices and the lossless matrix elements are finally combined to yield the scattering matrix S' for the lossy mixer. This final matrix has four rows and four columns for the simple mixers considered here, which is consistent with the observation that input power can be delivered to four possible loads. From a microwave point of view, the mixer is equivalent to a lossless junction which interconnects four terminated transmission lines, such as shown in Fig. 1.

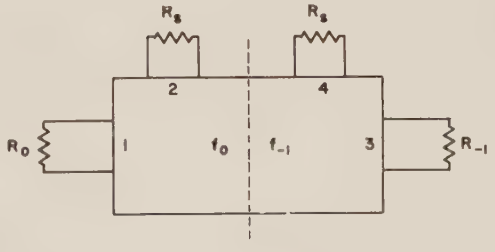


Fig. 1—"Junction" model of a parametric amplifier.

[10] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—part I. General energy relations," Proc. IRE, vol. 44, pp. 904–913; July, 1956.

[11] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," M.I.T. Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, N. Y., vol. 8, pp. 146–148, 288–289; 1948.

[12] C. R. Boyd, "A generalized approach to the evaluation of N-frequency parametric mixers," *Proc. Natl. Electronics Conf.*, vol. 16, pp. 472–479; October, 1960.

For the parametric amplifier, the gain, K, is given by the squared magnitude of the reflection coefficient at the signal port; *i.e.*,

$$K = |S_{11}|^2. \tag{8}$$

The excess temperature,[13] T_e, of a parametric amplifier is the equivalent noise temperature associated with the gain-normalized noise in the amplifier output wave which is not a consequence of noise incident on the amplifier in an input wave. This excess noise is the result of a black-body radiation into the "junction" of Fig. 1 from the loads R_s at f_0 and f_{-1}, and from R_{-1} at f_{-1}, which is subsequently scattered to the signal port. Since the noise waves from these sources are uncorrelated, they are summed power-wise, with the result

$$T_e = \frac{|S_{12}|^2 T_s + |S_{13}|^2 T_{-1} + |S_{14}|^2 T_s}{|S_{11}|^2}. \tag{9}$$

It is readily shown that the most favorable band-center condition for the parametric amplifier with respect to noise figure is that in which the signal and lower sideband circuits are resonated independently. For this circumstance, the steps outlined above are readily carried out; the parametric amplifier gain is found to be

$$K = \frac{\left[(R_0 - R_s)(R_{-1} + R_s) + \dfrac{\sigma_1^2}{\omega_0\omega_{-1}}\right]^2}{\left[(R_0 + R_s)(R_{-1} + R_s) - \dfrac{\sigma_1^2}{\omega_0\omega_{-1}}\right]^2}, \tag{10}$$

and its excess temperature at band-center is given by

$$T_e = 4\frac{R_0 R_s (R_{-1} + R_s)^2 T_s + \dfrac{\omega_{-1}}{\omega_0} R_0 \dfrac{\sigma_1^2}{\omega_0\omega_{-1}} [R_s T_s + R_{-1} T_{-1}]}{\left[(R_0 - R_s)(R_{-1} + R_s) + \dfrac{\sigma_1^2}{\omega_0\omega_{-1}}\right]^2}. \tag{11}$$

III. Load Considerations in Lossy Parametric Amplifiers

Experience indicates that the lossy parametric amplifier is capable of arbitrarily large gains, and the load condition which makes this possible is evident in (10) above. The denominator of this gain expression can be made to vanish by satisfying the requirement

$$(R_0 + R_s)(R_{-1} + R_s) = \frac{\sigma_1^2}{\omega_0\omega_{-1}}. \tag{12}$$

The form of this load constraint implies an upper frequency limit of operation for the lossy parametric amplifier. For example, consider the case of the quasi-degenerate mode amplifier, where ω_0 and ω_{-1} are essentially equal. In order to obtain maximum stable gain with this configuration, it is necessary to adjust the values of R_0, R_{-1} and σ_1 so that the following relationship holds at the particular frequency of operation f_0:

$$f_0 = \frac{\sigma_1}{2\pi\sqrt{(R_0 + R_s)(R_{-1} + R_s)}}. \tag{13}$$

If an attempt is made to increase f_0, the value of σ_1 must be increased, or the mixer loads reduced. In a practical diode capacitor, the largest value $\sigma_{1\,\max}$ to which σ_1 can be raised is limited by forward conduction and reverse breakdown. The external loads R_0 and R_{-1} can, of course, be set equal to zero, so that the maximum frequency f_D for infinite gain (or border-line instability) is found to be

$$f_D = \frac{\sigma_{1\,\max}}{2\pi R_s}. \tag{14}$$

At this frequency, which Kim[14] has defined as the *dynamic cutoff frequency* of the diode, the complete decoupling of external loads means that only the thermal noise from R_s can be amplified.

This parameter, f_D, is a meaningful and useful tool for characterizing the performance of lossy parametric amplifiers. For the more general case of a nondegenerate parametric amplifier, the infinite-gain condition is given by (12). Since f_0 and f_{-1} are not degenerate, the equivalent cutoff situation is, by analogy,

$$\sqrt{f_0 f_{-1}} = f_D. \tag{15}$$

That is, when the geometric mean of the signal and sideband frequencies is equal to the dynamic cutoff frequency, the external loading at both frequencies must be reduced to zero in order to obtain infinite gain.

The introduction of the concept of the dynamic cutoff frequency, f_D, enables the load constraints to be specified concisely. Thus, the maximum external loading for infinite regenerative gain occurs for the degenerate mode case when

$$(R_s + R_0)(R_s + R_{-1}) = R_s^2\left(\frac{f_D}{f_0}\right)^2. \tag{16}$$

The equivalent condition for the nondegenerate para-

[13] W. W. Mumford has pointed out to the author that the quantity "excess temperature" as used in this paper is synonymous with the IRE standard definition "effective input noise temperature."

[14] C. S. Kim, "Four-Terminal Equivalent Circuits of Parametric Diodes," G. E. Co. Rept. R59E2S-15, p. 19; February, 1959.

metric amplifier is

$$(R_s + R_0)(R_s + R_{-1}) \doteq R_s^2 \frac{f_D^2}{f_0 f_{-1}} . \tag{17}$$

The band-center gain expression of (10) may then be modified to incorporate the parameter f_D:

$$K = \left[\frac{-\frac{R_0}{R_s} + \left(1 - \frac{f_D^2}{f_0 f_{-1}}\right)}{\frac{R_0}{R_s} + \left(1 - \frac{f_D^2}{f_0 f_{-1}}\right)\left(1 + \frac{R_{-1}}{R_s}\right)} \right]^2 . \tag{18}$$

The particular manner of writing (18) demonstrates that the effect of idler loading on the band-center gain of the parametric amplifier is equivalent to reducing the quality of the diode. A *reduced dynamic cutoff frequency*, f_R, may be defined such that

$$f_R \equiv \frac{f_D}{\sqrt{1 + \frac{R_{-1}}{R_s}}} , \tag{19}$$

in which case the amplifier gain expression becomes,

$$K = \left[\frac{-\frac{R_0}{R_s} + \left(1 - \frac{f_R^2}{f_0 f_{-1}}\right)}{\frac{R_0}{R_s} + \left(1 - \frac{f_R^2}{f_0 f_{-1}}\right)} \right]^2 . \tag{20}$$

A family of curves showing the way in which gain, loading, and frequency parameters are related has been plotted in Fig. 2, using (20). For the special case of the quasi-degenerate mode amplifier, the source and idler loads are equal; then (18) may be written as

$$K = \left[\frac{1 - \left(\frac{R_0}{R_s}\right)^2 - \left(\frac{f_D}{f_0}\right)^2}{\left(1 + \frac{R_0}{R_s}\right)^2 - \left(\frac{f_D}{f_0}\right)^2} \right]^2 . \tag{21}$$

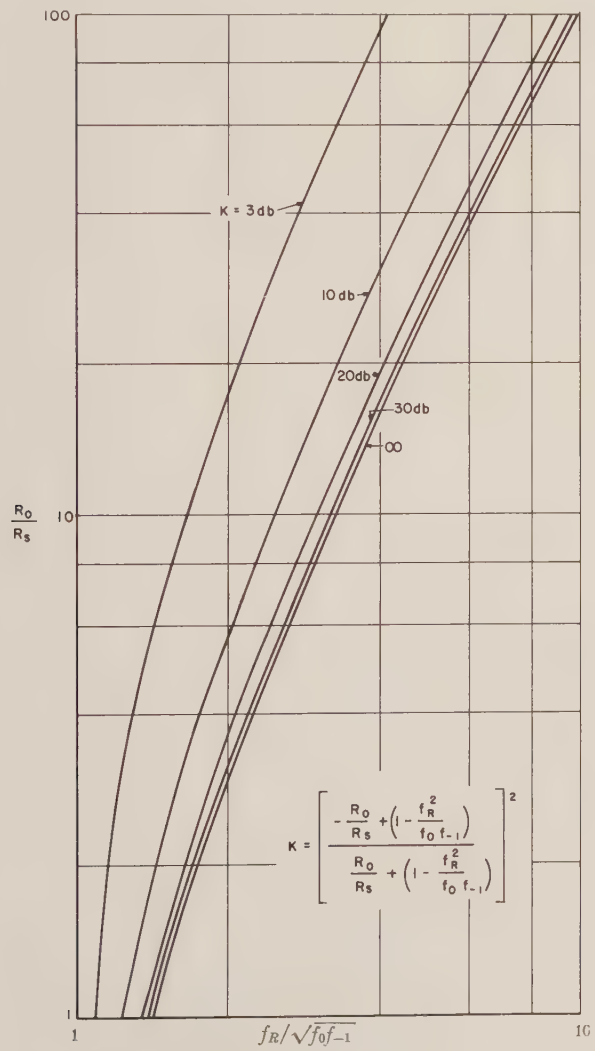


Fig. 2—Parametric amplifier constant midband gain characteristics.

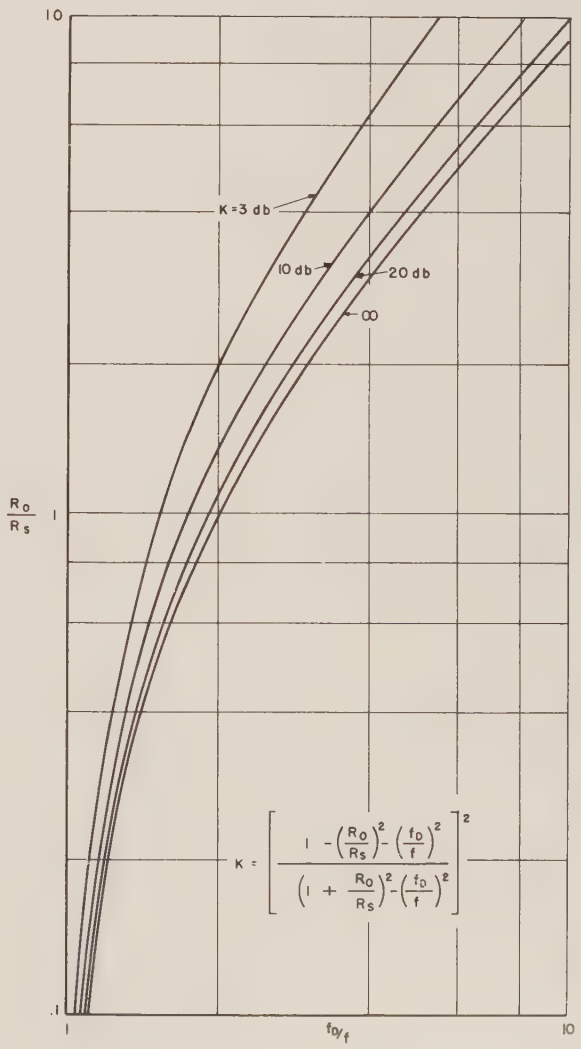


Fig. 3—Degenerate amplifier constant midband gain characteristics.

The curves of Fig. 3 show the manner in which gain, loading, and frequency are related according to (21).

IV. Optimizing Lossy Parametric Amplifier Excess Temperature

The constraint imposed on the external loading of a parametric amplifier by the series loss resistance R_s in conjunction with the dynamic elastance σ_1 sets a definite lower limit to the excess temperature which can be obtained under high-gain conditions for a given signal frequency and a given diode dynamic cutoff frequency. The first step in demonstrating this limitation is to incorporate the high-gain condition of (12) into the amplifier excess temperature expression, (11). This leads to the high-gain, band center excess temperature,

$$T_e \approx \frac{R_s T_s}{R_0} + \frac{f_0}{f_{-1}}\left(\frac{R_0 + R_s}{R_{-1} + R_s}\right)\left(\frac{R_s T_s}{R_0} + \frac{R_{-1} T_{-1}}{R_0}\right). \quad (22)$$

Of course, R_{-1} and R_0 are related by (17) as follows:

$$R_{-1} = R_s\left[\frac{f_D^2}{f_0 f_{-1}(1 + R_0/R_s)} - 1\right]. \quad (23)$$

After substitution for R_{-1} in (22), it is possible to find the input loading which minimizes the excess temperature. In the case of any parametric amplifier operated at uniform ambient temperature such that $T_{-1} = T_s$, the optimum ratio of R_0/R_s is the maximum value $f_D^2/f_0 f_{-1} - 1$, indicating that the idler load should be zero.

Using this optimum value of R_0/R_s for the parametric amplifier, and setting R_{-1} to zero, the excess temperature expression of (22) becomes

$$T_e \approx T_s\left[\frac{\left(\frac{f_D}{f_{-1}}\right)^2 + 1}{\left(\frac{f_D}{f_0}\right)\left(\frac{f_D}{f_{-1}}\right) - 1}\right]. \quad (24)$$

Note that for f_0 and f_{-1}, both well below dynamic cutoff, $f_D/f_0 \gg 1$, $f_D/f_{-1} \gg 1$, and (24) approximates the lossless case form

$$T_e \approx T_s \frac{f_0}{f_{-1}}. \quad (25)$$

The excess temperature as given by (24) has been plotted in Fig. 4 for some specific signal-to-dynamic cutoff frequency ratios. These curves illustrate that for a given signal frequency and diode dynamic cutoff frequency, there exists a broad minimum for the excess temperature of a parametric amplifier without external idler loading. The optimum idler frequency for absolute minimum excess temperature is found by the use of calculus to be

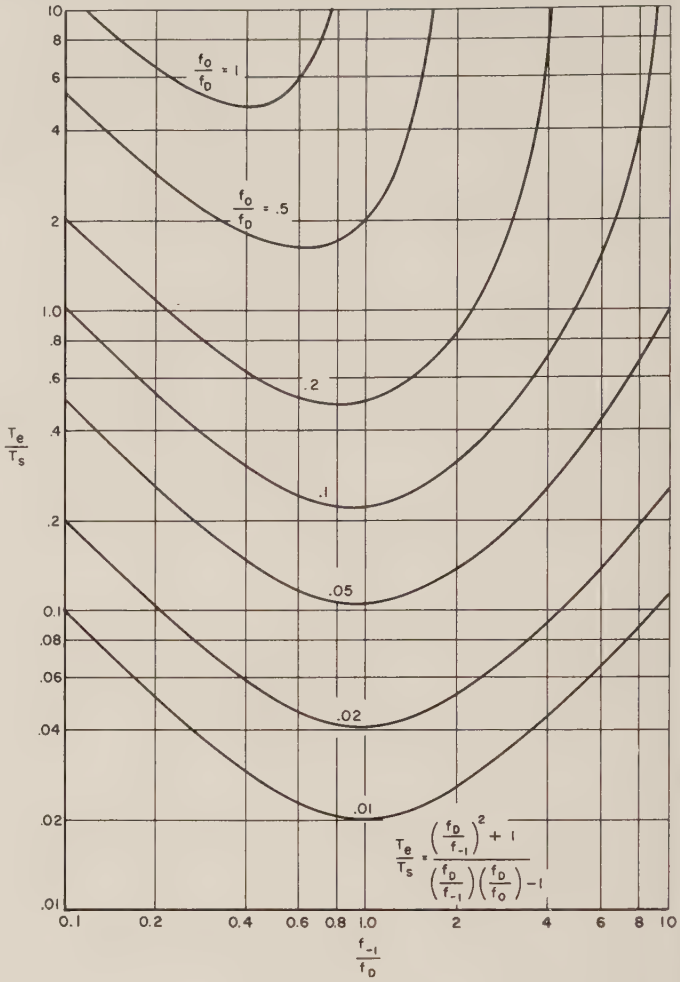


Fig. 4—Minimum excess temperature for a parametric amplifier without external idler loading.

$$f_{-1\,\text{opt.}} = f_D\left(\sqrt{\left(\frac{f_0}{f_D}\right)^2 + 1} - \frac{f_0}{f_D}\right). \quad (26)$$

The normalized optimum pump frequency is therefore given by

$$\frac{f_{p\,\text{opt.}}}{f_D} = \frac{f_{-1\,\text{opt.}}}{f_D} + \frac{f_0}{f_D} = \sqrt{\left(\frac{f_0}{f_D}\right)^2 + 1}. \quad (27)$$

This optimum pump frequency and the minimum excess temperature associated with it are in agreement with the values given by Greene and Sard.[3] Specifically, the optimum normalized excess temperature is

$$\left(\frac{T_e}{T_s}\right)_{\text{opt.}} = 2\left(\frac{f_0}{f_D}\right)^2\sqrt{\left[1 + \left(\frac{f_D}{f_0}\right)^2\right]}. \quad (28)$$

The relationships of (27) and (28) will be used in further discussion, and accordingly have been plotted in Fig. 5.

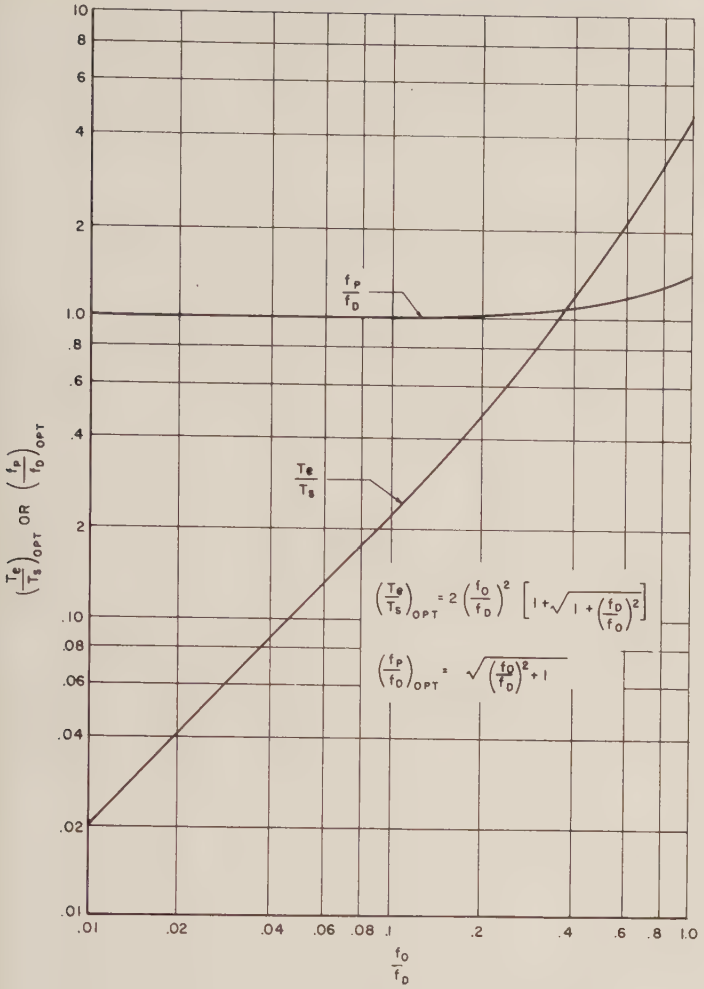


Fig. 5—Optimum excess temperature and pump frequency (Greene and Sard).

V. Relation Between Dynamic and Static Cutoff Frequencies

The above analysis based on the quantity f_D is, of course, meaningful for practical purposes only if f_D can be measured or at least correlated to commonly measured quantities. Although the value of f_D can in principle be computed from a capacity-voltage curve and an R_s measurement for a particular diode by performing a Fourier analysis to extract $\sigma_{1\ max}$, this method is too cumbersome to be of use in routine determination of diode quality. The resonance method described by Houlding,[6] on the other hand, appears to be a relatively convenient characterization technique, but suffers from the fact that it does not take into consideration the shape of the capacity-voltage curve and is not really representative of the actual operating conditions for the diode.

As a corollary program to an investigation of X-band parametric amplifier performance, a series of noise figure measurements were made on an X-band amplifier configuration to determine if a simple relation existed between the static cutoff, f_c, measured by Houlding's method and the dynamic cutoff, f_D, postulated in the previous analysis. The vehicle chosen for measurement was a quasi-degenerate mode amplifier, operated with a signal frequency at the lower edge of the X band; *i.e.*, at 8200 Mc. The use of the quasi-degenerate mode was based on considerations of simplicity regarding both the circuit manipulation and the noise figure relationship. In the case of the latter, analysis shows that the double-sideband noise figure (the quantity given by a noise figure measurement) is given by

$$F = \frac{1}{1-\left(\frac{f}{f_D}\right)^2} \tag{29}$$

under uniform ambient conditions. This elementary relationship is plotted in Fig. 6.

Degenerate-mode amplification was obtained at about 8200 Mc with a number of diffused-junction mesa-type diodes, fabricated from silicon and germanium. The noise figure of the amplifier was measured by means of a calibrated automatic noise figure indicator, with the gas-discharge noise source decoupled approximately 50 db from the amplifier through the use of two cascaded ferrite isolators. A conventional balanced mixer was used as the second stage, and the amplifier gain was adjusted to 15 db, on a single-sideband basis, for all measurements. The relatively low amplifier gain and the high decoupling of gas-discharge noise source from the amplifier were precautionary measures taken to assure that slight changes in impedance between the "on" and "off" conditions of the source did not result in appreciable amplifier gain modulation, and subsequent erroneous noise figure readings.

Fig. 7 shows the adjusted noise figure data obtained from these measurements after accounting for insertion loss in the measuring circuit and second-stage noise contributions. The best noise figure achieved with each diode is plotted against the operating frequency normalized by the f_c of that diode. In addition, a family of computed noise figure curves for constant f_c/f_D ratios has been plotted over the experimental data. A qualitative correspondence between the experimental and analytical behavior is evident, and an f_c/f_D ratio of about 10 is indicated.

VI. Application of the Noise Theory to the Design of a Low-Noise Preamplifier

With a complete first-order noise theory at his disposal, the designer of low-noise parametric amplifier circuits is in a position to make *a priori* judgments concerning the quality of diode and the pump frequency needed to satisfy his requirements. To illustrate the problems involved in choosing a diode and a pump frequency, we shall consider the design of a hypothetical

narrow-band low-noise amplifier for an L-band receiver operating at an arbitrarily chosen frequency of 1500 Mc.

In a typical situation where broad-band performance is not required, the designer is faced with a number of problems:

1) A particular performance level must be attained or exceeded. This invariably can be translated into a maximum allowable excess temperature for the amplifier.
2) Since high-cutoff diodes are associated with a correspondingly high price (a proportionality constant between one and two dollars per kMc is roughly correct at present), it is desirable to use the minimum quality diode, allowing for some safety margin, which meets the excess temperature specifications.
3) Some degree of variability must be accommodated with regard to diode dynamic cutoff frequency, even though the static cutoff frequencies of all diodes used are equal. Since the static cutoff frequencies of even carefully selected diodes are likely to vary ± 5 per cent, an allowance of at least ± 10 per cent in effective cutoff frequency should be made.

It is evident from the noise theory that the pump frequency and diode quality cannot be chosen independently. If a diode of insufficient quality is used, it may be impossible to meet the performance requirements at any pump frequency. If too high a pump frequency is used, a diode of higher than minimum cutoff frequency may be necessary to keep within the excess temperature maximum. If too low a pump frequency is used, even the best diode may not be good enough. What is needed is some criterion by which the minimum diode quality and optimum pump frequency may be chosen simultaneously.

Such a criterion is offered by the optimum pump frequency-minimum excess temperature curves presented in Fig. 5. Using these curves we may choose as our pump frequency that which produces the maximum allowable excess temperature using the lowest possible quality of diode. The design-center cutoff value for the diodes used would then be set about 10 per cent (more or less, depending on experience) higher than the marginal value to allow for variability of diodes. While this con-

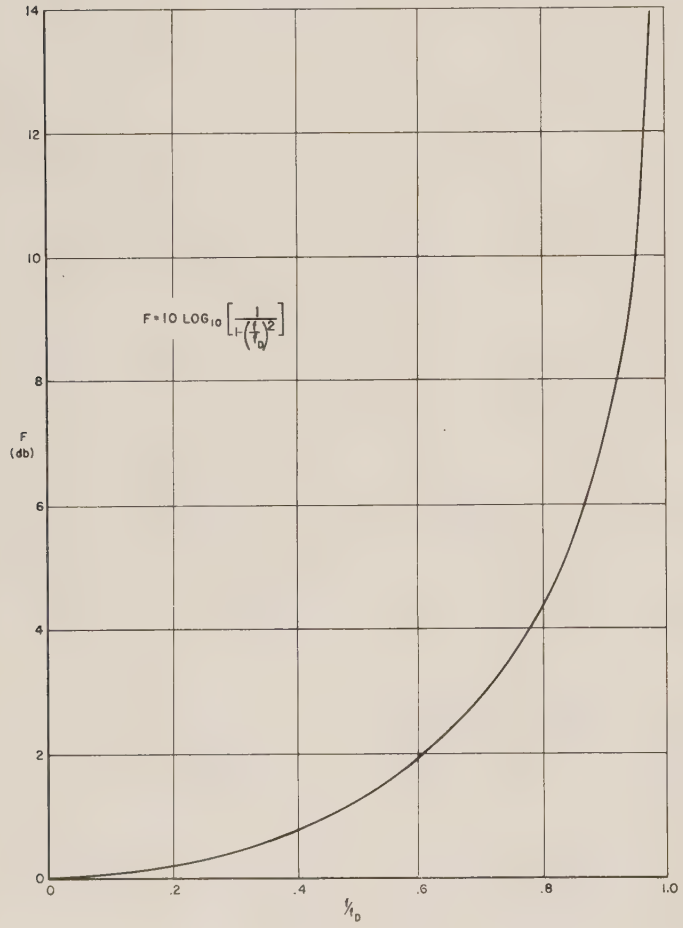


Fig. 6—Minimum noise figure, degenerate mode amplifier.

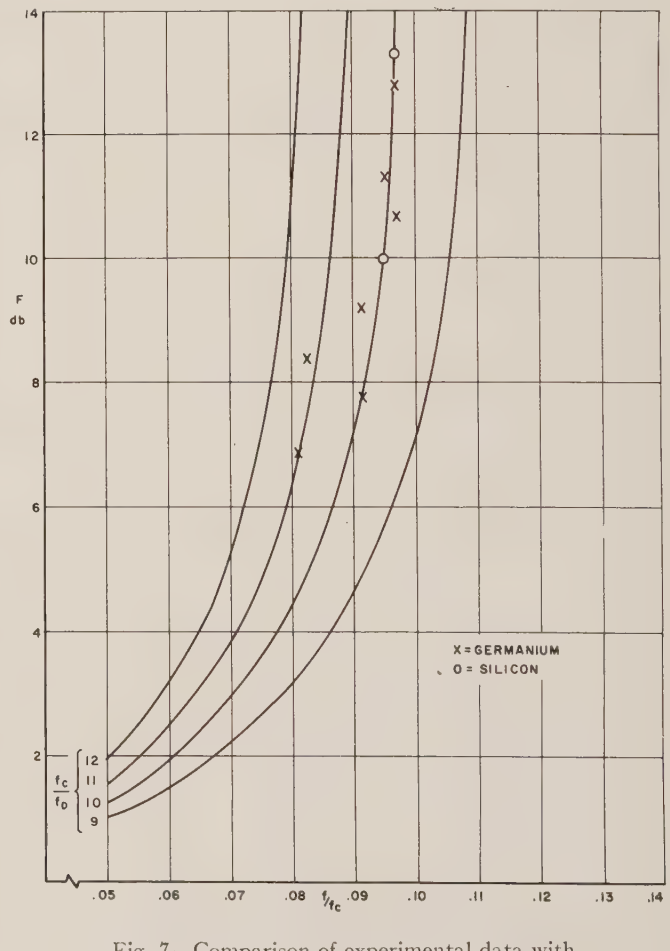


Fig. 7—Comparison of experimental data with computed performance.

servative design procedure will not squeeze the best possible performance out of design-center diodes, it provides the greatest margin for diode variability, and on that basis is probably a good compromise approach.

Suppose that our hypothetical 1500 Mc amplifier is required to operate with a high-gain, band-center excess temperature which does not exceed 114°K, under room-ambient conditions where $T_s = 290$°K. From Fig. 5, we see that the marginal performance condition is obtained for $f_D = 6f_0$, and $f_p = 1.014\ f_D$. Using the ratio $f_c/f_D = 10$, we choose 90 kMc as the lower limit for diode static cutoff, and prescribe a pump frequency of 9130 Mc. Because of the 10 per cent safety margin, a design-center static cutoff frequency of 100 kMc is established. Now the variation of excess temperature with diode cutoff can be computed for the prescribed pump frequency and compared with the pump-optimized values as in Fig. 8. It is seen that the conservative design procedure has cost only 1°K of excess temperature for design-center diodes compared with the pump-optimized design-center case.

As a matter of fact, the curves of Fig. 8 indicate that even if diodes of much higher cutoff frequency are used in our L-band amplifier, we will realize the greater portion of the total potential improvement without requiring a change of pump frequency. The reason for this behavior, of course, is the previously noted fact that the excess temperature minimum is very broad for the spectral arrangements of most interest from a low-noise performance point of view.

Because of this broad minimum of excess temperature, we might rightly expect that an exchange of an inconvenient optimum pump frequency for a convenient nonoptimum pump frequency could be made by paying only a small price in terms of diode quality. Such a trade-off would be very desirable, for example, where solid-state harmonic-generator pump power supplies were to be used. In such cases the higher outputs available at lower pumping frequencies and the generally lower pump powers required at same, might easily be worth a moderate sacrifice to the extent of more expensive diodes. Analysis of the minimum quality diode required to achieve the necessary 114°K excess temperature for our hypothetical L-band amplifier shows that our feelings are indeed justified. The curve of Fig. 9 reveals that the pump frequency can be dropped by as much as 25 per cent if we are willing (and able) to move the minimum diode quality up to 100 kMc.

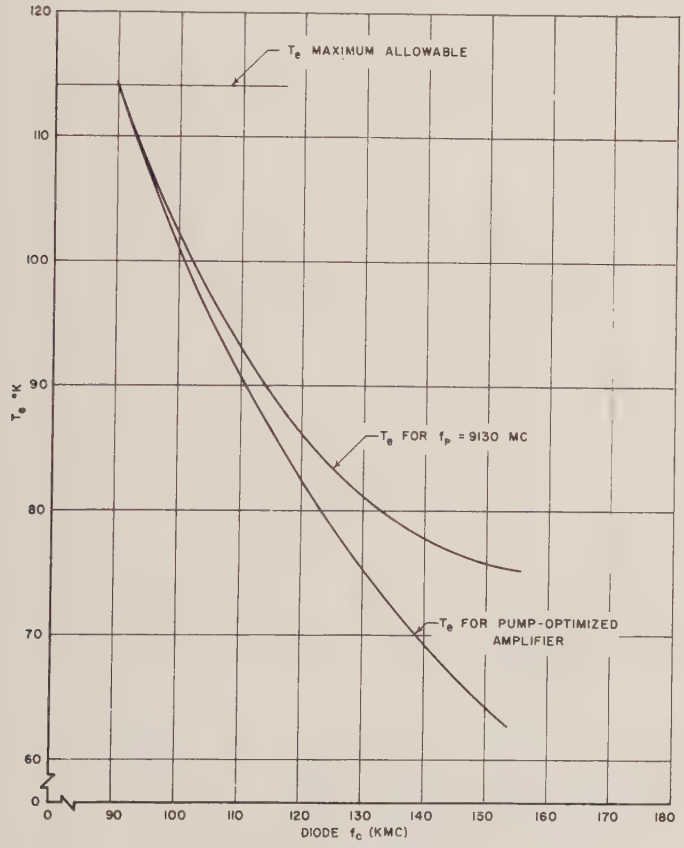


Fig. 8—Effects of improved diode quality on excess temperature.

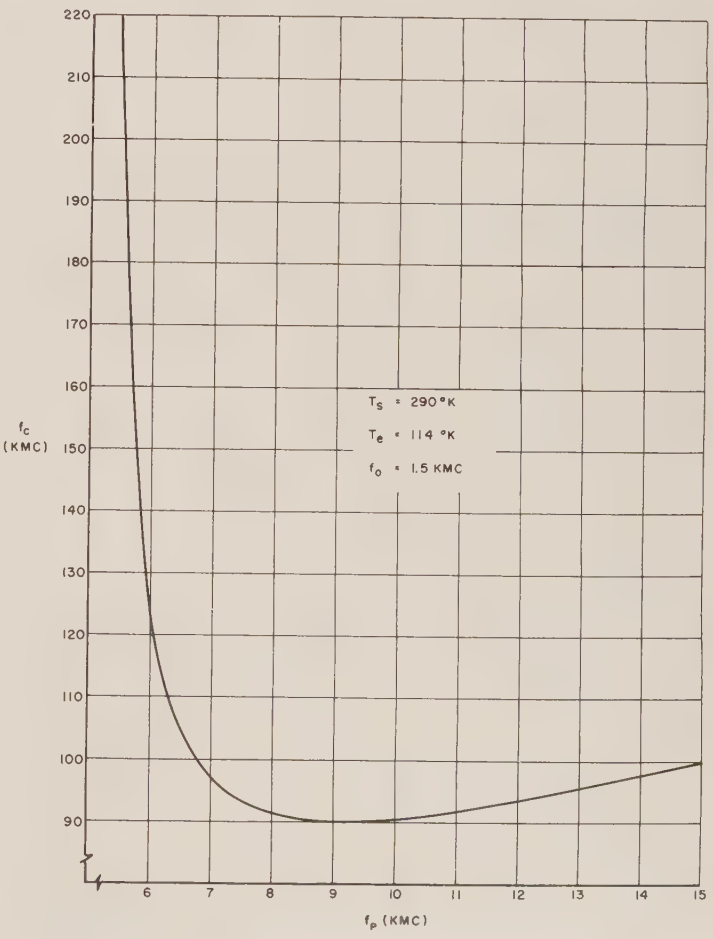


Fig. 9—Minimum diode quality as a function of pump frequency.

VII. Conclusions

A relatively complete first-order noise theory of parametric amplifiers is presently available to designers of such circuits. This theory is based on a rather simplified model for the diode capacitor, and probably should be extended in the near future to account for voltage-variable series resistance of the diode, and for the effects of varying barrier resistance. The voltage-variable series resistance case is likely to become more important when diodes having very thin base regions become available.

The existence of an optimum pumping frequency for a given signal frequency and given diode quality certainly should influence the thinking of the designer of parametric amplifiers, although perhaps not only in the most obvious manner. That is, the optimum pump frequency which provides the maximum allowable excess temperature with the minimum quality of diode should be viewed as the upper limit on pump frequency for that particular application, with a trade-off between higher minimum diode quality and lower pump frequency possible within limits. The use of a lower pump frequency is almost always desirable because of ease of circuit design, potential bandwidth improvement, and possible use of solid-state pump sources.

These conclusions, of course, have been drawn from analysis based on amplifiers which are not constrained with respect to bandwidth. In cases where bandwidth optimization is mandatory, it may be necessary to sacrifice diode quality in order to meet an excess temperature specification.[3]

Acknowledgment

It is a pleasure to acknowledge the encouragement of H. H. Grimm, of the General Electric Company Electronics Laboratory, Syracuse, N. Y., without which this paper would not have been prepared.

Steerable Array Radars*

FRANK C. OGG, Jr.†, senior member, ire

Summary—**The general characteristics of radars using large planar steerable array antennas are discussed. The need for an amplifier for each element is shown, and the tolerances and stability requirements for the amplifiers are discussed. Array geometry, pattern formation and gain, mutual coupling, and beam-steering techniques are summarized. Element minimization and signal-processing techniques are analyzed.**

I. Introduction

MECHANICAL scanning of directive antennas has been almost universally employed in existing radars. A horn-and-reflector allows only a limited scanning pattern and has a vulnerable mechanical structure, but has low initial cost and negligible maintenance.

Weapon performance has increased rapidly in recent years, bringing with it increased radar requirements. Among these are shock resistance, rapid volumetric scanning with high angular resolution, random beam positioning for multiple target tracking, very high radiated power, very large apertures and pulse-to-pulse tunability. One possible solution is the steerable array, a large fixed array of radiating elements with the beam steered by the introduction of a phase gradient across the aperture.

The best-known early example of a steerable array is the MUSA, which was built by Bell Telephone Laboratories before World War II for reception of transatlantic radio telephone signals.[1,2] A set of three independently steerable beams was formed from a linear array of rhombics, each beam being a horizontal fan with about one degree vertical beamwidth.

During World War II, steerable array radars using mechanical phasing were developed. A series-fed linear array (Fig. 1) was developed for a Naval fire-control radar.[3,4] A possible alternate configuration is the parallel-fed array (Fig. 2). A linear slotted-waveguide array was designed for mounting in the leading edge of an aircraft wing, with the guide wavelength changed by moving one wall of the waveguide to scan the beam.[5]

Since World War II, electrical phase shifting has been investigated to permit random beam positioning

* Received by the PGMIL, January 23, 1961. The research herein was supported by the USAF through Rome Air Dev. Ctr. of the Air Res. and Dev. Command. Preparation of the material for publication was supported through WWRNGW of the Wright Air Dev. Div.

† Rad. Lab., The Johns Hopkins University, Baltimore, Md. Formerly with the Bendix Corp., Towson, Md.

[1] H. T. Friis and C. B. Feldman, "A multiple-unit steerable antenna for short-wave reception," Proc. IRE, vol. 25, pp. 841–917; July, 1937.

[2] F. A. Polkinghorn, "A single-sideband MUSA receiving system for commercial operation on transatlantic radio telephone circuits," Proc. IRE, vol. 28, pp. 157–170; April, 1940.

[3] H. D. Friis and W. D. Lewis, "Radar antennas," *Bell Sys. Tech. J.*, vol. 26, pp. 219–317; April, 1947. (See especially p. 300.)

[4] G. C. Southworth, "Principles and Applications of Waveguide Transmission," D. Van Nostrand, Inc., New York, N. Y.; 1950.

[5] Friis and Lewis, *op. cit.*, see especially p. 315.

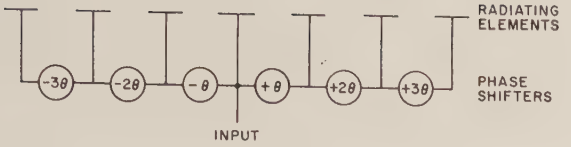


Fig. 1—Linear array with series feeding.

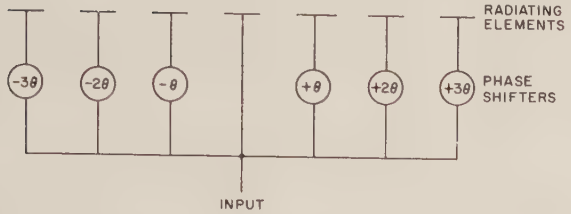


Fig. 2—Linear array with parallel feeding.

and more rapid scanning. The slotted-waveguide array may be scanned without moving the waveguide wall by variation of the frequency, leading to the folded-waveguide frequency-scanner (Fig. 3). Several wavelengths of guide between slots are used to reduce the tuning range for wide-angle scanning to 5 to 10 per cent. The development of ferrite phase-shifters permits electrical scanning of the configurations of Fig. 1 and Fig. 2.[6–9] Two-dimensional electrical scanning can be obtained by the "phase-phase" scanner, an obvious modification of either Fig. 1 or Fig. 2, or by the "phase-frequency" scanner (Fig. 4).

Neither the "phase-phase" nor the "phase-frequency" scanner fully satisfies modern requirements. Neither can form multiple steerable receiving beams without losses due to signal-splitting before amplification. Neither can radiate very high power, since present ferrite phase-shifters are limited to comparatively low peak powers. The frequency scanner is not tunable, in the sense that frequency at each beam position is fixed. In parallel-fed arrays of both types, recirculation of reflected power often spoils the pattern, since multiple reflections through the phase-shifters produce coherent radiation at other angles than the beam position. Isolators or circulators on each phase-shifter are usually necessary to maintain low sidelobes.[10]

[6] F. Reggia and E. G. Spencer, "A new technique in ferrite phase-shifting for beam scanning of microwave antennas," Proc. IRE, vol. 45, pp. 1510–1517; November, 1957.

[7] F. E. Goodwin and H. R. Senf, "Volumetric scanning of a radar with ferrite phase shifters," Proc. IRE, vol. 47, pp. 453–454; March, 1959.

[8] C. M. Johnson, "A ferrite phase shifter for the UHF region," IRE Trans. on Microwave Theory and Techniques, vol. MTT-7, pp. 27–31; January, 1959.

[9] D. D. King, C. M. Barrack, and C. M. Johnson, "Precise control of ferrite phase shifters," IRE Trans. on Microwave Theory and Techniques, vol. MTT-7, p. 229–232; April, 1959.

[10] Parallel-feeding is usually used in large arrays without amplifiers to avoid the losses of the series-fed type. Some measured patterns of parallel-fed arrays, with and without circulators, are given by S. J. Rabinowitz, "Array antennas with applications to radar," *Trans. of the University of Michigan Radar Symp.*, Ann Arbor, 1959, Willow Run Labs., The Univ. of Michigan; pp. 107–122.

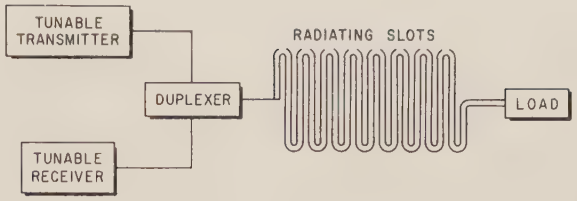


Fig. 3—Folded-waveguide frequency-scanning linear array.

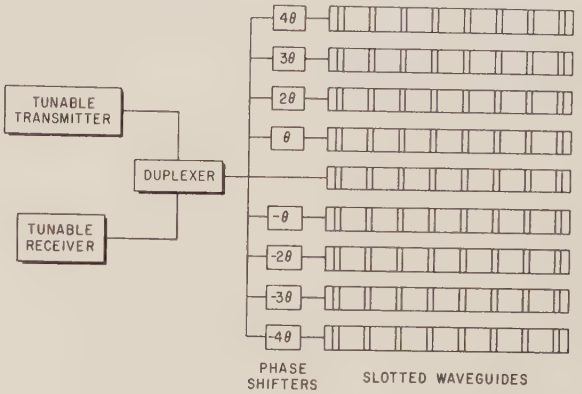


Fig. 4—Planar array with two-dimensional scanning using the "phase-frequency" principle.

These limitations have led to consideration of arrays with a transmitter and/or a receiver for each element. Very high radiated powers are possible with a multiplicity of transmitters of moderate size. The phase-shifters can be operated at low power levels. A receiver for each element allows formation of an unlimited number of beams for reception without loss of sensitivity. Tunability over the bandwidth of the amplifiers is possible, and the amplifiers provide isolation of reflected power.[11]

Many different configurations are possible with such an array. An arbitrary number of receiving beams can be formed, fixed or steerable as a stack of beams, or independently steerable. The transmitting array may floodlight a large angular region, illuminate the area covered by a steerable stack of beams, or illuminate separately several independently steerable beams. In the last type, the transmitting array may radiate a sequence of closely spaced pulses with the beam moved between pulses to illuminate the several receiving beams, thus obtaining a high data rate on many targets under simultaneous track. Both volumetric scanning and multiple-target tracking may be done simultaneously, eliminating the handover from a search radar to a separate tracking radar.

[11] The "phase-frequency" scanner (Fig. 4) can be modified by use of a transmitter and receiver on each slotted waveguide. Most of the objections to this system are then overcome. The array is still not tunable, and multiple receiving beams can be formed only in the "phase" dimension. The number of amplifiers required is reduced from N^2 to N and the tolerances are correspondingly tightened. Much of the analysis given in this paper is also applicable to a "phase-frequency" array with an amplifier for each row, but the principal subject of the paper is the "phase-phase" array with an amplifier for each element.

The choice of configuration for a specific application is beyond the scope of this paper, which is intended to be a general survey of possibilities and limitations.

II. Geometry of a Planar Array

Unlike the horn and reflector antenna, an array has a different pattern at each beam position. These pattern variations are rather complicated in the usual elevation-azimuth coordinates,[12] but they are greatly simplified when a more natural coordinate system is used.

The array factor of a planar array with equal element spacings is

$$F(\theta_1, \theta_2) = \sum_{m,n} a_{mn} \exp\left[2\pi m j d_1 \theta_1 + 2\pi n j d_2 \theta_2\right], \quad (1)$$

where d_1 and d_2 are the row and column spacings in wavelengths and $\theta_1 = \sin\alpha - \sin\alpha_0$, $\theta_2 = \sin\beta - \sin\beta_0$. In these equations, the direction cosines of the normal to the plane wave received by or radiated by the array are $\sin\alpha$ and $\sin\beta$. The beam maximum is at $\alpha = \alpha_0$, $\beta = \beta_0$. The phase distribution on the aperture is factorable in these coordinates, which simplifies beam steering.

The connection between the natural array coordinates and elevation-azimuth can be found when the array orientation is known. This connection is illustrated in Fig. 5 for an array tilted 40° from the vertical with scanning 45° from boresight.

The pattern variations with scanning are easily understood in the coordinates of Fig. 5. The relative positions of the various beams are preserved. With a vertical stack of beams, the stack remains vertical in these coordinates. When the stack is not on the vertical principal plane ($\beta_0 = 0$), it will not generally be vertical in elevation-azimuth coordinates; in the upper corners of Fig. 5 it will be almost horizontal. The same remark holds for the four-beam set of a monopulse system, as well as for the polarization of the radiated wavefront.

Changes in beam shape with scanning are also treated simply. The function $F(\theta_1, \theta_2)$ is fixed, and only the variation of θ_1 and θ_2 need be considered. For example, the half-power points of the beam in the two coordinates α, β are fixed values of θ_1 and θ_2. If

$$F(\theta_{10}, 0) = F(-\theta_{10}, 0) = F(0, \theta_{20}) = F(0, -\theta_{20}) = \frac{1}{\sqrt{2}} F(0, 0),$$

then the half-power points are located at $\sin\alpha - \sin\alpha_0 = \pm\theta_{10}$ and $\sin\beta - \sin\beta_0 = \pm\theta_{20}$, regardless of steering angle. The half-power points are thus dependent only on the half-power beamwidth with the beam at boresight and do not depend on any other characteristic of the array. It can be shown that the half-power beamwidth is given very accurately by the usual cosine-foreshortening approximation until the array approaches

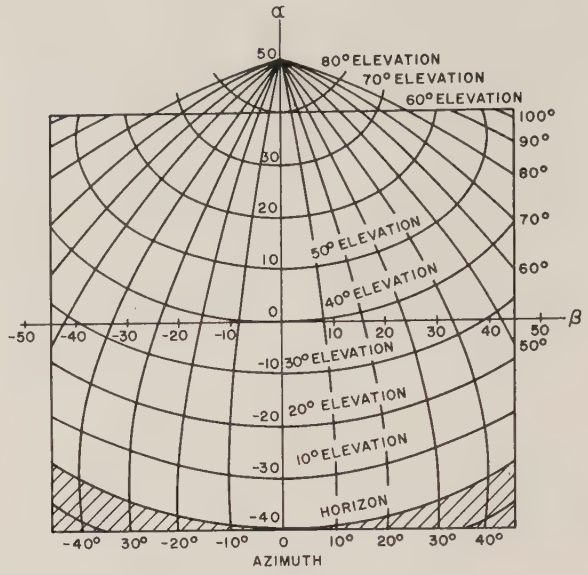


Fig. 5—Angular coverage of a planar array tilted 40° from vertical. Beam steering is theoretically possible from −45° to +45° in the horizontal steering angle β and from −40° to +40° in the vertical steering angle α. Elevation and azimuth contours are plotted on a rectangular grid of lines of constant α and β. The zenith is at $\alpha = 50°$, $\beta = 0°$.

end-fire.[13] An expression which is accurate at all angles is

$$\text{beamwidth} = \arcsin\left[\sin\alpha_0 + \sin\frac{\theta}{2}\right] - \arcsin\left[\sin\alpha_0 - \sin\frac{\theta}{2}\right], \quad (2)$$

where θ is the half-power beamwidth at boresight. This is the array factor beamwidth, which in small arrays or near end-fire may be modified by the element pattern. Other beam characteristics, such as asymmetry, may be treated similarly.

The array is operated entirely in the natural array coordinates. The fact that the resulting scan patterns and beam shapes may be rather unusual when viewed in elevation-azimuth is usually irrelevant. The radar data can be converted if necessary to any desired coordinates, but it is convenient to perform beam interpolation, scanning-pattern generation, tracking, etc., in the natural coordinates.

The actual position of the beam with respect to the plane of the array is shown in Fig. 6. The beam generated by a linear array lies on a cone about the axis of the array, since this cone is the locus of points with the same direction cosine. In a planar array with a factorable phase distribution, the beam lies on the intersection of the two cones shown. When both natural array angles are at 45°, for example, the beam lies in the plane of the array. The theoretical coverage of Fig. 5 must be modified by superposition of the element factor upon it.

[12] A detailed treatment of array geometry is given by W. H. von Aulock, "Properties of phased arrays," Proc. IRE, vol. 48, pp. 1715–1727; October, 1960.

[13] R. W. Bickmore, "A note on the effective aperture of electrically scanned arrays," IRE Trans. on Antennas and Propagation, vol. AP-6, pp. 194–196; April, 1958.

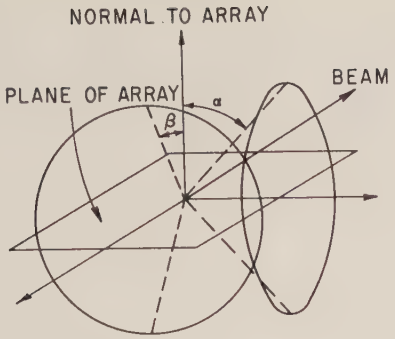


Fig. 6—Beam position with respect to the plane of the array. The phase distribution on the aperture is factorable. The phase distribution in the angle α produces a cone of radiated energy about one axis, while the distribution in β produces a cone about the other axis. The beam is the intersection line of the two cones. The second intersection line (below the plane of the array) is usually suppressed by the element factor. A linear array produces one cone, with part of the cone usually suppressed by the element factor.

III. Pattern and Gain of an Array

If I_{mn} is the actual current in the m,nth element, and if the isolated element factor is $E(\alpha, \beta)$, the radiated pattern of the array is proportional to

$$F(\alpha,\beta) = E(\alpha,\beta) \sum_{m,n} I_{mn} \exp\left[2\pi jmd_1(\sin\alpha - \sin\alpha_0) + 2\pi jnd_2(\sin\beta - \sin\beta_0)\right]. \quad (3)$$

Eq. (3) assumes that the elements are far enough apart so that the adjacent elements do not modify the distribution of current in the element.

The gain of the array is equal to

$$G(\alpha,\beta) = \frac{F(\alpha,\beta)\cdot F^*(\alpha,\beta)}{P_0},$$

where P_0 is the total radiated power. The total radiated power may be computed by integration of the Poynting vector over a sufficiently large sphere Ω. The result is

$$P_0 = \frac{1}{4\pi}\int\int_\Omega F(\alpha,\beta)\cdot F^*(\alpha,\beta)dS = \sum_{mn}\sum_{pq} I_{mn}\cdot I_{pq}^* \exp\left[-2\pi j(m-p)d_1\sin\alpha_0 - 2\pi j(n-q)d_2\sin\beta_0\right]R_{mnpq}, \quad (4)$$

where

$$R_{mnpq} = \frac{1}{4\pi}\int\int_\Omega E(\alpha,\beta)\cdot E^*(\alpha,\beta)\exp\left[2\pi j(m-p)d_1\sin\alpha + 2\pi j(n-q)d_2\sin\beta\right]dS \quad (5)$$

is proportional to the mutual radiation resistance. If the element factor is normalized by

$$\frac{1}{4\pi}\int\int_\Omega E(\alpha,\beta)\cdot E^*(\alpha,\beta)dS = 1,$$

then R_{mnpq} is the mutual radiation resistance normalized by the self-radiation resistance. The normalized mutual radiation resistance given by (5) is dependent only on the isolated element pattern and on the spacing. In the case of a theoretical isotropic element, the calculation is simple and gives the result

$$R_{mnpq} = \frac{\sin 2\pi R}{2\pi R},$$

where $R^2 = d_1^2(m-p)^2 + d_2^2(n-q)^2$ is the square of the separation of the two elements in wavelengths.

The main beam gain of the array, assuming that the excitation currents are held constant, is generally dependent on beam position, and is given by

$$G(\alpha_0,\beta_0) = \frac{E(\alpha_0,\beta_0)\cdot E^*(\alpha_0,\beta_0)\left(\sum_{mn} I_{mn}\right)^2}{P_0(\alpha_0,\beta_0)}.$$

The exact evaluation of the gain can be carried out either by direct integration of the pattern or by (4) and (5). A calculation for a uniformly illuminated planar array is reported in the literature and shows that the gain varies approximately as the change in beamwidths would indicate.[14]

In a linear array,

$$P_0 = \sum_m \sum_p I_m\cdot I_p^* \exp\left[-2\pi j(m-p)d_1\sin\alpha_0\right]R_{mp}, \quad (6)$$

where

$$R_{mp} = \frac{1}{4\pi}\int\int_\Omega E(\alpha,\beta)\cdot E^*(\alpha,\beta)\cdot\exp\left[2\pi j(m-p)d_1\sin\alpha\right]dS, \quad (7)$$

which are specializations of (4) and (5). When the elements are isotropic,

$$R_{mp} = \frac{\sin 2\pi d_1(m-p)}{2\pi d_1(m-p)}.$$

When the element spacing is a half-wavelength, all mutual radiation resistances are zero and the gain remains constant with scan angle. A geometrical explanation is given by Fig. 6. The beam lies along a cone and with isotropic elements includes the whole cone. As the beamwidth broadens, the arc of the cone decreases.

Exact calculation of the gain of an array is quite complicated. In general, the usual estimates from the beamwidth are approximately correct if the conical beam shape of a linear array is taken into account.

If a directive element is used, the energy cannot be accounted for when the main beam is steered into a null of the element pattern. The main beam disappears or is

[14] R. K. Thomas and M. J. King, "Gain of large scanned arrays," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 635–636; November, 1960.

greatly reduced in gain, and the energy in the main beam apparently is not radiated. It has been suggested that this energy is somehow stored in the antenna structure and reflected back to the transmitters, creating a very large SWR at each transmitter. Attempts to observe this effect experimentally have generally failed, and it is not clear how to account for the energy.[15]

On reception, the pattern is formed in a beam-forming network (Section VIII) with tapering after amplification. If the signal components are equal before tapering, then they add coherently after tapering, while the uncorrelated receiver noise components add incoherently. An output signal-to-noise ratio of

$$\left(\frac{S}{N}\right)_{\text{out}} = \left(\frac{S}{N}\right)_{\text{in}} \frac{(\sum a_n)^2}{\sum a_n^2} \tag{8}$$

is obtained, where a_n are the taper coefficients.

If the noise is local oscillator noise rather than receiver noise, it will be coherent rather than incoherent. When the beam is at boresight, the local oscillator noise will add coherently. At other beam positions, it is reduced by the array factor, since it will be combined after phasing. The magnitude of the oscillator noise is increased in proportion to the signal. An array will therefore have a blind spot at boresight shaped like the receiving beam unless the local oscillator noise is sufficiently suppressed.

IV. Antenna Coupling Effects

The pattern and gain of an array with known currents in its elements are discussed in Section III. The currents actually present are modified by antenna coupling effects and are therefore dependent on the currents in the other elements and on beam position.

An exact calculation of the actual currents is possible in principle as a large set of coupled-network equations if the mutual impedances are calculated or measured.[16] In practice, however, the order of this set of linear equations is the number of elements in the array, and the calculation is impractical except in small arrays. A large array can be regarded approximately as infinite in extent, and the coupling effects can be obtained for an infinite uniformly illuminated array without excessive difficulty. This approximation can be used to obtain information about coupling effects in receiving arrays, which are always uniformly illuminated, or in uniform transmitting arrays.

The principal coupling effects are a change in the effective element pattern, gain variations in scanning due to mismatches between antenna and amplifier, and variations in the loads driven by the transmitters. In finite or nonuniform arrays, the taper is changed during scanning by unequal coupling effects on the various elements.

In an infinite uniformly illuminated array, each element is affected by coupling in exactly the same way. The taper is unchanged, and there are no pattern variations during scanning except those discussed in Section II.[17]

The element factor is modified by passive reflections from the other elements present in the array. The element factor used in pattern calculations is the pattern of an element embedded in an infinite array of similar elements. With half-wave spacing, the effective element factor is usually broader than the isolated element pattern, which is often useful in wide-angle scanning systems.[18]

The gain variations are due to changes in the antenna impedance during scanning which prevent matching transmitters and receivers to the antennas except at one beam position. These variations are usually small (1 db or less in a dipole array) except near end-fire.[19,20] The variable load presented to the transmitters causes a large reflected current which is often significant. Standing-wave ratios of about 2:1 are encountered in uniformly illuminated dipole arrays scanned 45° from boresight.

In a finite array, the elements near the edge are affected differently. In a large array, edge effects are usually negligible. It is possible, but usually unnecessary, to eliminate edge effects in receiving arrays with dummy elements around the edges of the aperture.

In a nonuniform array, an analysis based on an infinite uniform array is inapplicable. Very little is known about the effects of coupling on transmitting arrays with power tapering. When an irregular distribution similar to those of Section VI is used, the receiving array can be filled in with dummy elements to equalize the coupling effects. In a transmitting array of this kind, the effects are not equalized.

In Section III, the normalized mutual radiation resistance was shown to be dependent only on the isolated element pattern and on the spacing. This result appears to contradict experience, since a variety of elements with similar patterns are known to have very different measured couplings. These measurements are made by driving one element with a signal source and observing the

[15] This problem was pointed out by S. J. Rabinowitz.

[16] P. S. Carter, "Circuit relations in radiating systems and applications to antenna problems," Proc. IRE, vol. 20, pp. 1004–1041; June, 1932.

[17] J. Blass and S. J. Rabinowitz, "Mutual coupling in two-dimensional arrays," 1957 IRE WESCON Convention Record, pt. 1, pp. 134–150.

[18] The element factor is not always broadened and usually is rather irregular when embedded in an infinite array.

[19] Near end-fire these effects are much greater. See E. A. Blasi and R. S. Elliott, "Scanning antenna arrays of discrete elements," IRE Trans. on Antennas and Propagation, vol. AP-7, pp. 435–436; October, 1959.

[20] P. S. Carter, Jr., "Mutual impedance effects in large beam scanning arrays," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 276–285; May, 1960.

received signal at the terminals of the other element.[21] With elements having essentially the same pattern as a dipole, coupled power from 15 db to 30 db down has been observed at half-wave spacing. The discrepancy is due to the directive nature of the coupling in traveling-wave antennas, such as the helix or polyrod. Some measurements on polyrods are given by Southworth, showing that the coupling in one mode is about 30 db greater than in the other.[22] A forward-traveling wave is radiated in such an antenna while the reverse-traveling wave is reflected to the source. The radiated component produces the mutual radiation resistance given by (5) and the resulting pattern variation, while the reverse component is responsible for the standing-wave ratio at the transmitter. If care is taken in matching the element to free space, the reverse component can be considerably reduced. Isolation of the transmitters from antenna impedance variations can be obtained in this manner with strongly coupled elements. The conventional method for the measurement of mutual impedance measures only the reverse component.

V. Aperture Illumination Errors

Preservation of the pattern and gain of an array depends on approximate matching of the multiple parallel channels. The accuracy with which the channels can be matched determines the sidelobe level and gain loss, while the period of stability in which matching to the desired accuracy can be maintained determines the required frequency of realignment.

The exact nature of the channel drifts depends on the components used and their interconnections, but it can usually be assumed that they are independent and have the same distribution with mean zero in each channel. Calculation of the pattern statistics under these assumptions has been treated by several authors,[23] based on the fundamental work of Ruze.[24]

Ruze's analysis can be extended to include random amplifier failures.[25] The array factor including errors and failures is

$$F(\alpha,\beta) = \sum_{mn} I_{mn}\cdot F_{mn}\cdot(1+\Delta_{mn})\cdot\exp\left[j\delta_{mn}\right]\cdot\exp\left[j\phi_{mn}\right], \tag{9}$$

where I_{mn} and ϕ_{mn} are the design current and phase, F_{mn} is 0 or 1 as the element is failed or operating, Δ_{mn} is the fractional current error, and δ_{mn} is the phase error in radians, each for the m,nth element in the array. The power array factor is $P(\alpha,\beta)=F(\alpha,\beta)\cdot F^*(\alpha,\beta)$. If all sources of error are assumed independent, then the average power array factor is

$$\overline{P(\alpha,\beta)} = \left[(\overline{F})^2(\overline{\exp j\delta})^2\right]P_0(\alpha,\beta) + \left[(\overline{F^2})(1+\overline{\Delta^2}) - (\overline{F})^2(\overline{\exp j\delta})^2\right]\sum_{mn} I_{mn}^2, \tag{10}$$

where $P_0(\alpha,\beta)$ is the design-array factor. The distribution of amplitude error enters only as its variance $\overline{\Delta^2}$, and

$$\overline{F} = \overline{F^2} = 1 - \frac{F}{N},$$

where F is the number of failures and N is the number of elements. The phase error appears only as $\overline{\exp j\delta}$.

The calculated average pattern must be renormalized by $(1-F/N)^{-1}$, since the total power in the pattern is reduced by failures. The normalized average power array factor is then

$$\overline{P_N(\alpha,\beta)} = \left[\left(1-\frac{F}{N}\right)(\overline{\exp j\delta})^2\right]P_0(\alpha,\beta) + \left[1+\overline{\Delta^2}-\left(1-\frac{F}{N}\right)(\overline{\exp j\delta})^2\right]\cdot\sum_{mn} I_{mn}^2. \tag{11}$$

This expression shows that the effect of errors and failures consists of a gain reduction in the design-array factor with the energy removed from the main beam, redistributed around the pattern as an omnidirectional "noise level."

The gain loss is

$$\left(1-\frac{F}{N}\right)(\overline{\exp j\delta})^2 \tag{12}$$

and the rms noise level on the pattern is

$$\left[\frac{1+\overline{\Delta^2}-\left(1-\frac{F}{N}\right)(\overline{\exp j\delta})^2}{\left(1-\frac{F}{N}\right)(\overline{\exp j\delta})^2}\right]\frac{\sum_{mn} I_{mn}^2}{\left(\sum_{mn} I_{mn}\right)^2}, \tag{13}$$

when referred to the maximum amplitude of the beam.

If the phase errors are distributed normally with variance $\overline{\delta^2}$, then $(\overline{\exp j\delta})^2=\exp[-\overline{\delta^2}]$. If the channels are monitored and amplifiers which are out of a tolerance δ_0 are removed, then the distribution of phase

[21] E. Altshuler, "The measurement of self and mutual impedances," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 526–527; September, 1960

[22] Southworth, *op. cit.*, p. 439.

[23] A bibliography is given in R. S. Elliott, "Mechanical and electrical tolerances in two-dimensional scanning antenna arrays," IRE Trans. on Antennas and Propagation, vol. AP-6, pp. 114–119; January, 1958.

[24] J. Ruze, "The effect of aperture errors on the antenna radiation pattern," *Suppl. Nuovo Cimento* (Ital.), vol. 9, no. 3, pp. 364–380; 1952.

[25] This analysis is based on an unpublished Sanders Associates Memo. by S. J. O'Neil.

errors should become approximately uniform, and $\overline{(\exp j\delta)}^2 = (\sin \delta_0/\delta_0)^2$. For small errors distributed uniformly between tolerances Δ_0 and δ_0, and for small numbers of failures, (13) is approximately

$$\left\{\frac{\Delta_0^2}{3} + \frac{\delta_0^2}{3} + \frac{F}{N}\right\} \frac{\sum I_{mn}^2}{(\sum I_{mn})^2}. \tag{14}$$

The total pattern has a modified Rayleigh distribution regardless of the detailed distributions of error, and confidence levels for the distribution of sidelobe peaks can be computed when the design sidelobe level and the error statistics are given.[24] Although the assumptions of the above analysis are somewhat questionable, it is found in practice that the peak sidelobe level of an array can usually be predicted within about 1 db.

The gain loss is independent of array size and amplitude errors. Fig. 7 shows gain loss as a function of phase tolerance, assuming no failures. A phase tolerance of 50° holds the gain loss to 1 db, and a similar loss is produced by 20 per cent failures. If low sidelobes are unimportant, as is often true in transmitting arrays, only enough accuracy is needed to prevent excessive loss of gain.

Much tighter tolerances are required for low sidelobes. Tolerances for a given sidelobe level are proportional to $\sqrt{N}$, where N is the number of elements in the array. Fig. 8 shows the phase and amplitude tolerances required for an rms sidelobe level of −50 db, which will usually give a peak sidelobe level between −35 and −40 db.

Failure or removal for repair of a large fraction of the amplifiers has almost no effect on the pattern. Fig. 9 shows the fraction of elements which may be inoperative while a −30-db rms sidelobe level is maintained.

The required channel stability is a function of the number of channels and of the amount of maintenance which is possible. If an array radar with 10,000 transmitters and 10,000 receivers is to be maintained by a crew which can remove, repair, and realign a hundred amplifiers a day, then an average stability of about six months is necessary.

Channel matching and relative stability are complicated by the requirement for tunability of the radar. Since individual tuning of each amplifier is impossible, all RF components are broad band over the desired tuning range. Change of radiated frequency is then accomplished by oscillator tuning. Matching of antenna elements, duplexers and RF amplifiers over a large frequency band is considerably more difficult than matching at a single frequency.

From the equations given, the consequences of failure to maintain the correct aperture illumination can be computed. Methods of maintaining the aperture illumination over a long period of time fall into two classes: closed-loop stabilization and design of extremely stable conventional amplifiers. Interest in very-low sidelobes has led to investigations of phase-stabilization tech-

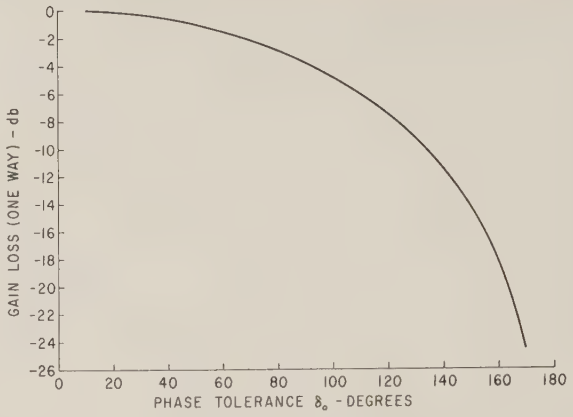


Fig. 7—Gain loss (one-way) as a function of phase tolerance, assuming a uniform distribution of phases between tolerance limits.

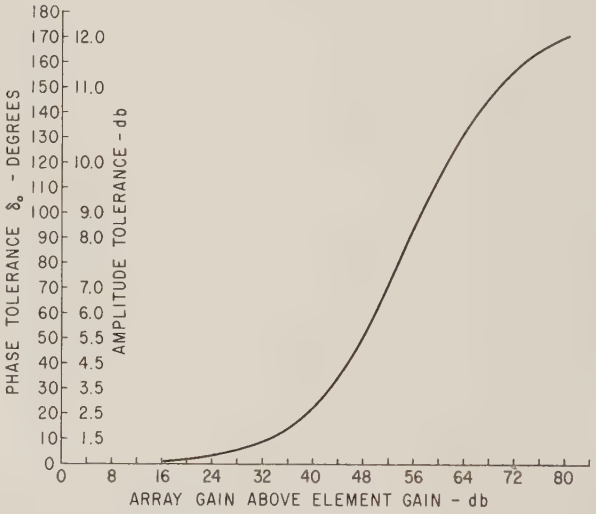


Fig. 8—Phase and amplitude tolerances required for −50-db rms sidelobe level in arrays of various sizes.

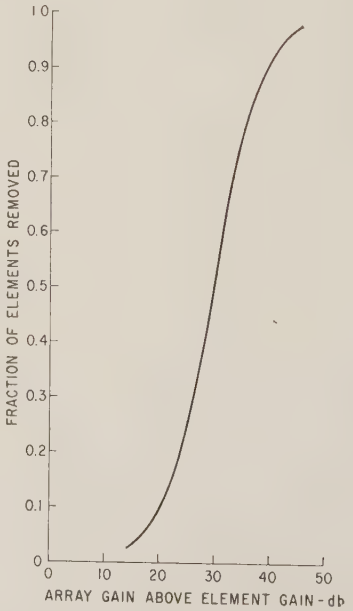


Fig. 9—Fraction of elements removable in a random manner without exceeding an rms sidelobe level of −30 db, in arrays of various sizes.

niques.[26–30] Stabilizers able to correct a half-cycle phase error in a few microseconds are available and could be used on each component subject to drift. Phase stabilization simplifies channel matching, since the channels are automatically matched in phase (although not in amplitude) over their bandwidth. It reduces the pattern distortions due to element coupling, since it is the actual current in the antenna element which is equalized in phase with the reference phase. The aperture distribution will therefore be correct in phase at all times, although the amplitude distribution may be distorted.

On the other hand, a considerable complication of the amplifier is necessary. Phase stabilization in itself is not sufficient to preserve low sidelobes; phase stabilization to one degree is useless unless amplitude variations can be held to about 0.1 db. The phase variations in the actual current will be large and rapid, even though the amplifier itself is stable, due to frequency changes and element coupling changes with beam steering. A stabilizer must therefore operate on each pulse separately and correct the phase within a small fraction of a pulse length.

For these reasons, a stable conventional amplifier is a more satisfactory solution when feasible; however, the feasibility of such amplifiers is beyond the scope of the present paper and is dependent on the array characteristics.

VI. Element Minimization

Minimization of the number of amplifiers required to achieve a given gain and/or angular resolution is of interest for practical and economic reasons. Fig. 9 suggests that random removal of elements leads to a pattern with fewer elements and the same resolution as a solid aperture distribution. In a sufficiently large array, a useful degree of directivity can be obtained with a very thin distribution of elements over the aperture. The sidelobe level in the resulting array is random and the peak sidelobe level will usually be about 10 to 15 db above the rms level given by (13). No grating lobes will appear.

Consideration of random element removal shows the fundamental limitations of thin aperture distributions. The gain is decreased in proportion to the number of elements removed. The gain removed from the main beam is redistributed around the pattern as a random omnidirectional component. In a full aperture nearly all of the radiated energy goes into the main beam, while in a thin distribution most of the energy appears in the sidelobe region. If 90 per cent of the elements in the aperture are removed, the main-beam gain is decreased by 10 db and about 90 per cent of the radiated energy is wasted in the sidelobe region rather than applied to target illumination.

These characteristics are not dependent on the random character of the distribution. Any removal procedure which preserves beamwidth and element spacing will lead to main-beam gain reduction and to an increase in the fraction of integrated gain in the sidelobe region. In extreme cases, such as elimination of alternate rows or columns, the energy removed from the main beam appears as a single large grating lobe. The average sidelobe level of a thin distribution is determined by the amount of energy removed from the main beam, which is dependent on the fraction of elements removed. The difference between elimination schemes is in the uniformity of the sidelobes. An ideal thin distribution would have regular sidelobes with equal amplitudes and the required average level. No synthesis methods for such distributions are known.

Thin distributions are therefore not applicable to transmitting arrays. If it is necessary to use a smaller number of transmitters, the only efficient solution is to use a smaller transmitting aperture and to obtain angular resolution with multiple simultaneous receiving beams. In receiving arrays, thin distributions can be used to maximize angular resolution for given gain. The required gain determines the number of receivers needed and these receiving elements can then be arranged in a thin distribution over a larger aperture. The limit on the size of the aperture (or the thinness of the distribution) is the acceptable sidelobe level.[31]

A substantial reduction in the number of amplifiers can be obtained by removing the tapering. A uniform array of unit elements (equal amplitude and no relative phase) is the most economical way to design both receiving and transmitting arrays. The only serious objection to uniform arrays is a high sidelobe level near the main beam.

In a transmitting array, power tapering is impractical for other reasons. It is extremely difficult to match amplifiers with a different number of stages over a band of frequencies. Maximum total power is generated by full power on each transmitter, and manufacturing and maintenance are simplified by identical transmitters.

Ideal distributions of unit elements are the optimum design for both transmitting and receiving arrays. Since no synthesis methods are known, some examples were

[26] "Phase Stabilization Techniques for Electrically Scanned Arrays," Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Tech. Rept. on Contract No. AF 30(602)-1862; June, 1959.

[27] J. R. Cummings, "A Differential Phase Stabilization System," M.S. thesis, Dept. of Elec. Engrg., Mass. Inst. Tech., Cambridge; May, 1957.

[28] W. P. Delaney, "A Phase Stabilization Technique for Pulsed UHF Power Amplifiers," M.S. thesis, Dept. of Elec. Engrg., Mass. Inst. Tech., Cambridge; June, 1959.

[29] K. W. Exworthy, "Two Systems for Accurate Microwave Phase Control," M.S. thesis, Dept. of Elec. Engrg., Mass. Inst. Tech., Cambridge; April, 1959.

[30] E. W. Markow, "Servo phase control shapes antenna pattern," *Electronics*, vol. 32, pp. 50–52; January 2, 1959.

[31] In some applications the antenna noise temperature may be increased by the increased gain in the sidelobe region.

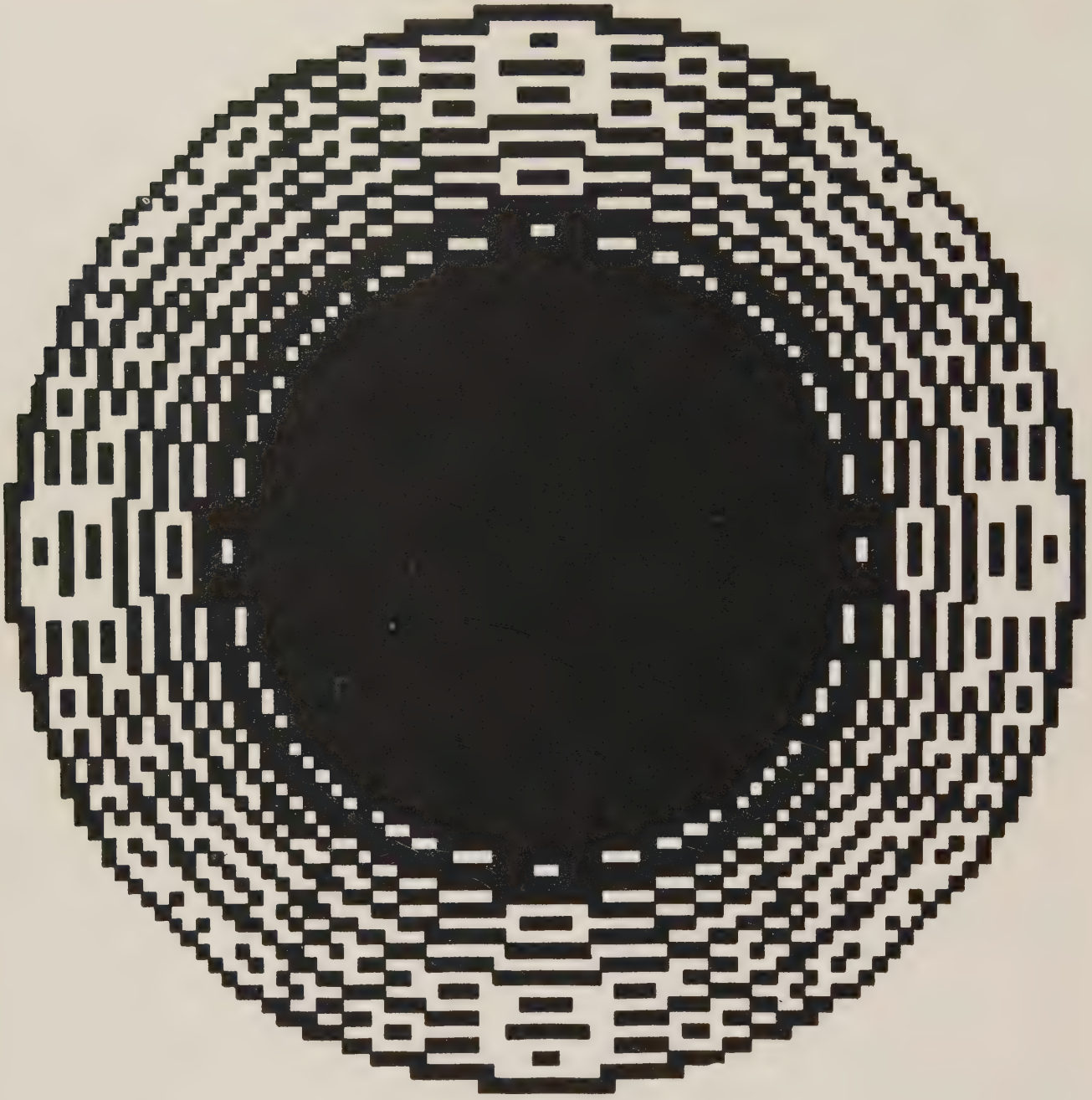


Fig. 10—Circular array of unit elements with 1.5° beamwidth and full gain. Element spacing approximately one-half wavelength with black spaces illuminated, white spaces omitted.

constructed empirically.[32] An ordinary amplitude distribution (usually Gaussian or Tchebycheff) was selected with the desired aperture, beamwidths and sidelobes. It was then normalized to make the sum of its taper coefficients equal to the number of unit elements desired in the distribution. For transmitting arrays, the number of unit elements was approximately equal to the gain above the element gain indicated by the beamwidths. For receiving arrays, the number of unit elements was a given fraction of this gain. Fractions from 20 per cent to 100 per cent were selected. The normalized amplitude taper then gives the average density of unit elements over the aperture, and the exact positions are found by experimentation.

Fig. 10 shows a circular array which forms a 1.5° beam with full gain from 3260 elements in an 80×80 aperture. Fig. 11 shows a circular array which forms a similar beam with about one fifth of full gain from 758 elements in a 78×78 aperture. Fig. 12 shows a rectangular array which forms a shaped fan-beam. In all three cases, the main beams are identical with those of the original amplitude tapers. The peak sidelobe levels are about -20 to -22 db. In the thin distribution (Fig. 11), the sidelobes remain high throughout the pattern, as in all thin distributions, but the patterns of the full-gain apertures have only a few sidelobes above -30 db. These empirical distributions are far from optimum, but are appreciably better than uni-

[32] These distributions were constructed by J. H. Best. A more detailed discussion of them will be presented at the next PGMIL Natl. Convention, Washington, D. C., June 26–28, 1961.

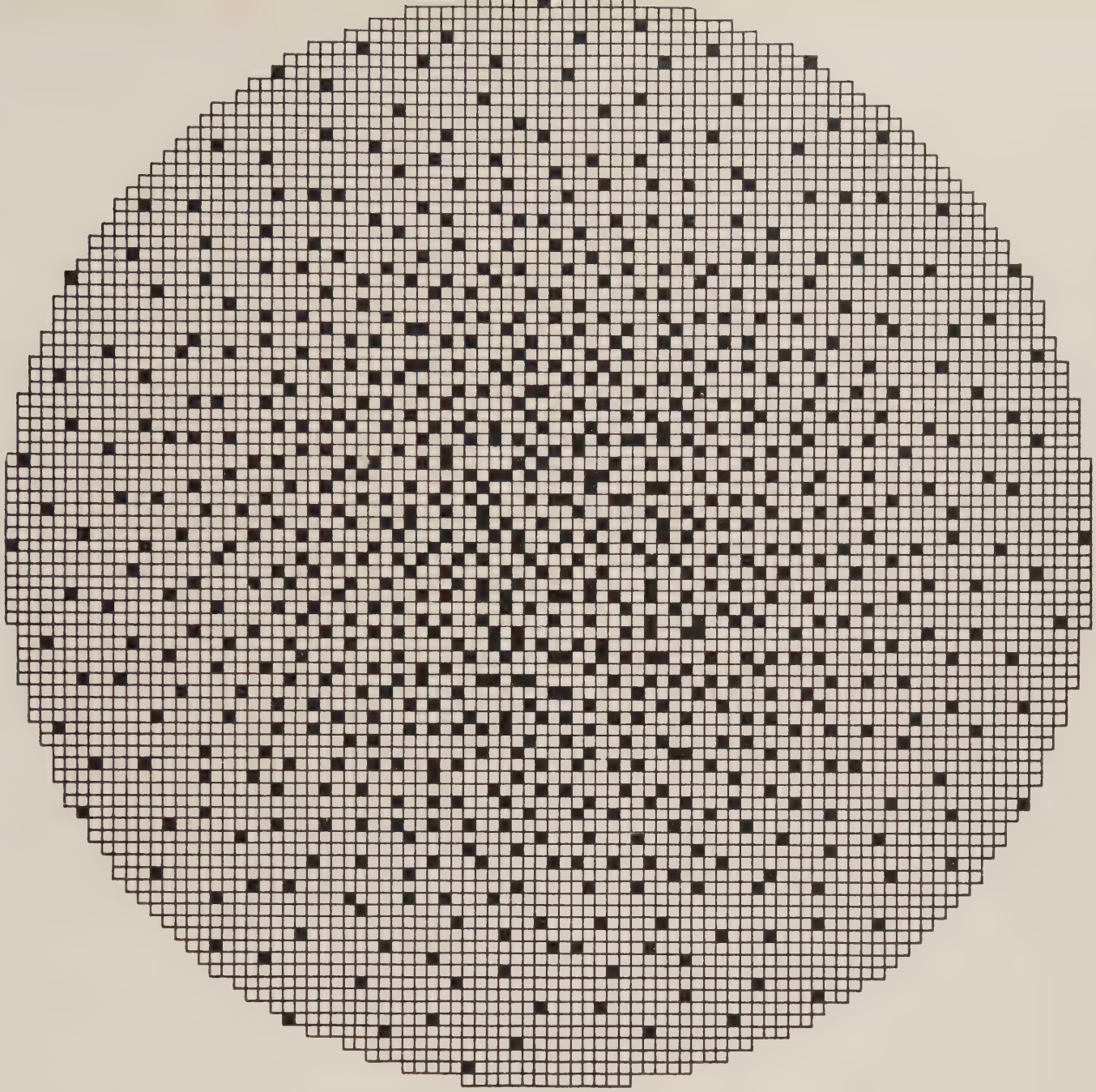


Fig. 11—Circular array of unit elements with 1.5° beamwidth and one-fifth of full gain. Element spacing approximately one-half wavelength with black spaces illuminated, white spaces omitted.

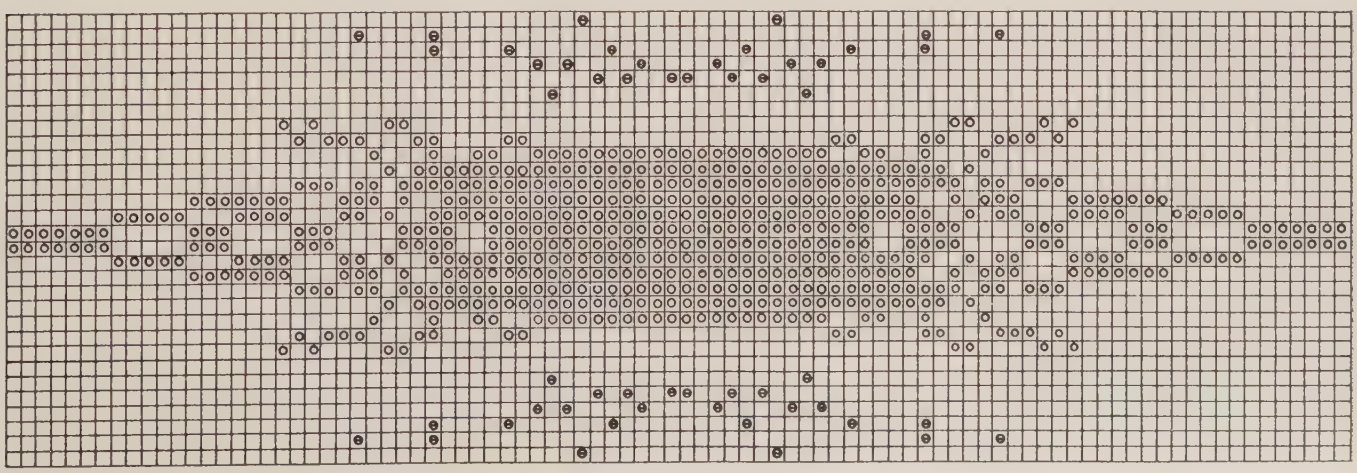


Fig. 12—Rectangular array of unit elements with shaped fan-beam and full gain. Element spacing approximately one-half wavelength. Circles indicate illuminated elements; circles with minus signs are illuminated 180° out of phase for beam shaping.

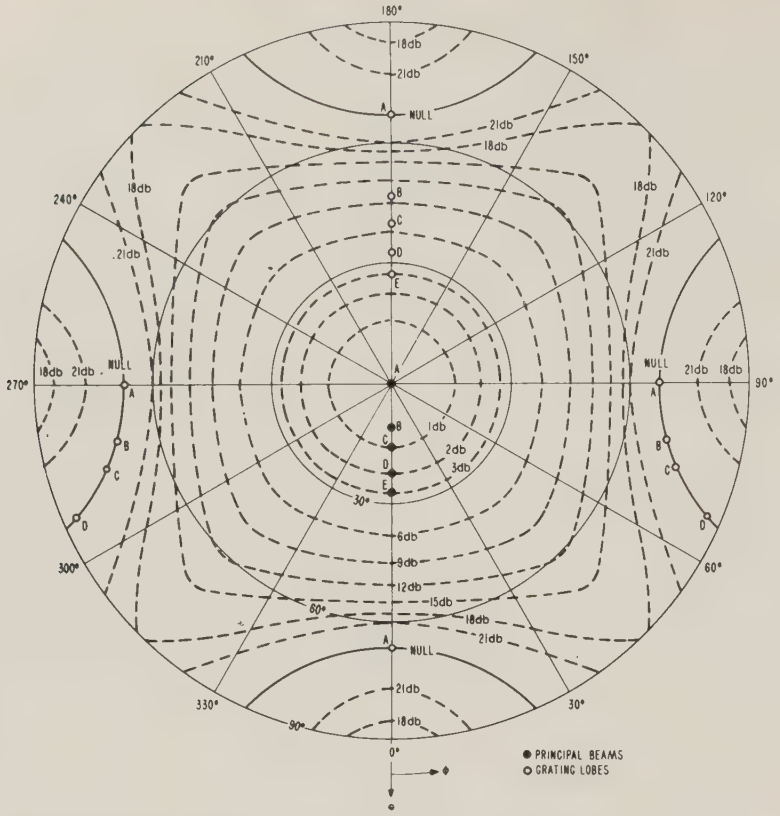


Fig. 13—Beam steering using a 2×2 subarray. The element factor of the subarray is superimposed as a contour plot. The points A, B, C, D, E, indicate various positions of the main array beam and its grating lobes.

form apertures. The thin distribution has irregular sidelobes, but is an improvement over a uniform aperture with random element elimination. The aperture area for the distributions of Figs. 10 and 11 is about half again as large as the aperture required for a comparable uniform distribution, adding to structure costs.

Another method of reducing the number of amplifiers is by the division of the array into subarrays with an RF phase-shifter on each element and an amplifier on each subarray. The subarray forms an "element" which is much more directive than the original element and must itself be steered for wide-angle scanning. The group of elements with RF phasing form a "steerable element" which is steered with the beam of the array.

In transmitting arrays which form a single beam, this technique is applicable, but is limited by the power-handling capability of the RF phase-shifter. In receiving arrays, a multiplicity of receiving beams is usually formed by a beam-forming network after the amplifiers. The effects of displacing the receiving beam from the direction in which the subarray is steered are shown in Fig. 13. When the array beam is at "A," the grating lobes fall into the nulls of the subarray pattern and are suppressed. When the beam is displaced slightly from the center of the subarray pattern, the grating lobe immediately appears. At "B," the displacement from the center of the subarray pattern is a small fraction of the subarray beamwidth, but the grating lobe is suppressed only 10 db. The smallest possible subarray (2×2) has been used in this calculation. With larger subarrays, the permissible beam offset would be smaller. The subarray technique is therefore not applicable when multiple beams are required unless the offset is very small. A stack of beams could be formed in elevation if each row were a linear subarray and if an amplifier were used in each row before summation of the row sums.

VII. Beam-Steering Techniques

In an array without amplifiers, beam steering is necessarily performed with RF phase-shifters. With an amplifier present on each element, a set of properly phased signals may be generated at any convenient frequency and then heterodyned to the desired frequency. Since lower-frequency circuitry is simpler and more stable than microwave components, an IF beam-steering system is appreciably simpler than a set of RF phase-shifters.

Since a linear phase gradient across the aperture is to be generated, a tapped delay line driven by a variable frequency oscillator suggests itself. The output frequency at the taps on the line will vary with beam position, since the beam is steered by changing the frequency in the line. To eliminate the resulting variation in the radiated frequency a dual mixing scheme is used, as shown in Fig. 14. This tapped delay line produces a linear phase gradient at a fixed frequency, since phase is preserved in a mixer and the frequency variation is eliminated. There are a number of possible variations of the system of Fig. 14.

Two such beam-steering systems are required for two-

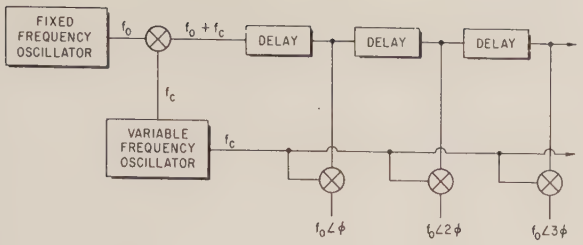


Fig. 14—IF beam-steering system with dual mixing and fixed output frequency.

dimensional beam positioning. The steering signal for the m,nth element is derived by mixing together the output of the mth tap of the horizontal-steering system and the nth tap of the vertical-steering system. The mixer output has a frequency which is the sum of the two input frequencies and a phase which is the sum of the two input phases.

The tolerances on the tap outputs are very tight. Each tap controls the phase of an entire row or column of the array and its phase must be held to a few degrees even in a very large array. The equations of Section V are applicable if the tapped delay line is considered to be a linear array with a number of elements equal to the number of rows or columns in the array. These tolerances are more easily met at lower frequencies, but the delay line becomes longer and its size and weight increase.

The IF beam-steering system imposes a lower bound on range resolution which may be important in some applications.[33] When a very short pulse is received from a target not at boresight, the signal will enter the amplifiers on one side of the array before reaching those on the other side. The pulse will be delayed equally in each channel and, at the output of the beam-forming network, will have a ramp-like leading edge and trailing edge of length $D \sin \theta/c$, where D is the aperture diameter, c is the speed of light, and θ is the angle off boresight. If the pulse-length is less than this quantity, which is the transit time across the aperture, the pulse will never reach full amplitude. In most arrays, the transit time is a small fraction of a microsecond and the effect is not significant. If very high range resolution is wanted, then beam steering must be done with time-delays rather than phase-shifts in order to compensate the delay across the aperture. Ferrite phase-shifters will approximately resemble time-delays, and so a series-fed ferrite-phased array may be made to have range resolution less than the transit time.[34]

IF beam-steering systems also produce a beam shift with a change of radiated frequency. This effect must be taken into account in setting the variable-frequency oscillator which controls the beam position.

[33] L. R. Dausin, K. E. Niebuhr, and N. J. Nilsson, "The effects of wide-band signals on radar antenna design," 1959 IRE WESCON Convention Record, pt. 1, pp. 40–48.

[34] C. M. Johnson, "Bandwidth of ferrite phase shifters for phased array and direction-finding use," Proc. IRE, vol. 47, p. 1665, September, 1959.

VIII. Beam-Forming Networks

The formation of a monopulse or stacked-beam configuration of receiving beams requires that the elementary signals, after amplification and conversion to IF, be divided and then combined with the proper phase-shifts to form beams offset from boresight. In this manner, an IF beam-forming network can be used to form a set of beams which can be steered as a group.[35]

In a planar array, the rows are usually summed first and then the row sums are combined as a column. If M beams are formed in each row-combining network and N beams are formed in the column-combining network, then a set of $M \times N$ beams which are fixed with respect to each other are formed. For convenience, only a linear array is considered here.

A beam squinted at an angle θ from boresight is created by forming the sum

$$\sum_n I_n \exp [2\pi jnd \sin \theta]. \tag{15}$$

The simplest type of network uses a parallel RC combination to provide the phase-shift; that is, the sum

$$\sum_n I_n [R_n^{-1} + j\omega C_n] \tag{16}$$

is formed with

$$R_n^{-1} = \pm \mid \cos (2\pi nd \sin \theta) \mid$$
$$\omega_0 C_n = \pm \mid \sin (2\pi nd \sin \theta) \mid ,$$

where ω_0 is the IF used in the network. The appropriate sign is obtained by inversion in a double-ended amplifier. A part of a simplified form of an RC network is shown in Fig. 15.

This network provides the desired output when the signal is a sine wave at the design frequency ω_0, but not in the presence of an unpredictable Doppler shift or a broad-band signal.

The network corresponding to (16) can be separated into two networks, a resistive network $\sum_n I_n [\cos 2\pi nd \sin \theta]$ and a reactive network $j \sum_n I_n \sin [2\pi nd \sin \theta]$. Only the reactive network distorts the pattern. The patterns due to these two networks considered separately are shown in Fig. 16. The cosine-amplitude taper produces beams at θ and $-\theta$ with the same polarity, while the sine taper produces the same two beams with opposite polarity. When the two networks produce the same output, the two spurious beams at $-\theta$ cancel, while the two beams at θ add.

The Doppler shift of a sine wave of frequency ω produces a network output of

$$I = I_0 + j \sum_n \left(\frac{\omega - \omega_0}{\omega_0} \right) \omega_0 C_n I_n. \tag{17}$$

[35] Beam-forming networks of this type were originated and developed by Sanders Assoc., Inc. Section VIII is based on some unpublished Sanders Assoc. memoranda and on an unpublished Bendix report by J. C. Nolen.

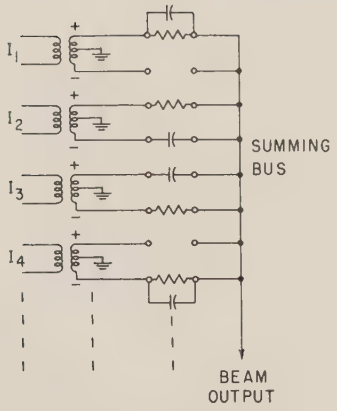


Fig. 15—Simplified RC beam-forming network forming one squinted beam.

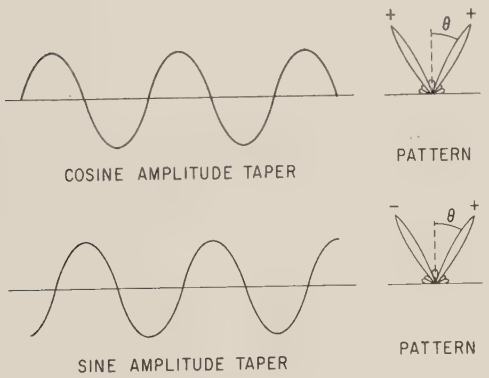


Fig. 16—Patterns of resistive and reactive parts of beam-forming network, illustrating analysis of spurious responses.

The term I_0 is given by (15) and produces the correct (single-lobed) pattern, while the second term of (17) produces a double-lobed spurious response. The magnitude of the spurious response in the beam (at θ) is ω/ω_0-1, which is usually negligible compared with the correct pattern, while the spurious response in the sidelobe region (at $-\theta$) is $1-\omega/\omega_0$. The spurious beam amplitude, referred to the main beam, is therefore $\omega_0-\omega/2\omega_0$, and the Doppler shift must be less than 2 per cent of the carrier frequency ω_0 if the spurious response is to be 40 db below the main beam. This effect is a serious limitation if rapidly moving targets are present, if the IF is low, or if the radiated frequency is high.

Distortion of a waveform $f(t)$ can be treated similarly. If the frequency spectrum of the waveform is $F(\omega)$, then the pattern distortion is given for each frequency component by the second term of (17). The frequency spectrum of the waveform at IF is $F(\omega-\omega_0)$ and the spectrum of the network output is $(\omega-\omega_0/\omega_0)$ $F(\omega-\omega_0)$. The waveform of the output is

$$\hat{f}(t) = \int_{-\infty}^{+\infty}\left(\frac{\omega-\omega_0}{\omega_0}\right)F(\omega-\omega_0)e^{-j\omega t}d\omega$$

$$= -\frac{j}{\omega_0}\frac{df}{dt}e^{j\omega_0 t}. \qquad (18)$$

The waveform which appears in the main beam is $f(t)-j/2\omega_0 f'(t)$, and that which appears in the spurious beam is $j/2\omega_0 f'(t)$. The waveform after passage through the beam-forming network thus has a distortion which depends on the pulse shape. This distortion is chiefly important in tracking systems, since it distorts the leading edge of the pulse.

These pattern distortions can be reduced by compensating the capacitances with inductances, and the response of the network can be analyzed in the same manner. A better solution is to form the real and imaginary parts of (15) in separate resistive networks and to combine their outputs in a broad-band quadrature phase-shifter, as in Fig. 17. The pattern distortions of this network, which are essentially due to the quadrature phase-shifter, can be analyzed by the same method. In this way, a beam-forming network which will handle a wide-bandwidth signal without pattern or waveform distortion can be designed.

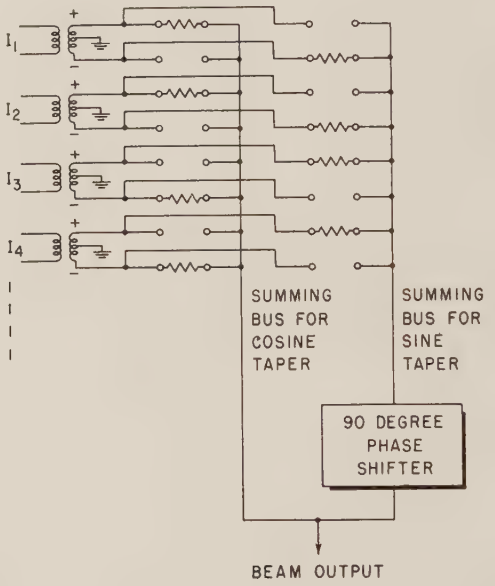


Fig. 17—Broad-band beam-forming network forming one squinted beam.

IX. Receiving Arrays with Clipping

With a sufficiently strong signal, infinite clipping in each receiver normalizes each elementary signal without changing its phase and thus normalizes the received pattern as a whole without changing its polar plot or directivity properties. An antenna without sidelobes but with the desired main beam can be obtained by setting a threshold above the sidelobe level of the normalized pattern.

If the nonlinear amplitude characteristic of the receiver is $F(a)$, and if the receiver preserves the phase of the signal, then the received pattern in a linear array will be

$$\sum_n a_n F(I_n)\exp\left[-jn\phi_0\right], \qquad (19)$$

where a_n is the taper coefficient (applied after the nonlinear operation), I_n is the nth elementary signal and ϕ_0 is the phase gradient for beam steering. If only a single signal is present, and if noise is neglected, $I_n = I_0 \exp(jn\phi)$ and $F(I_n) = F(I_0) \exp(jn\phi)$ due to phase preservation in the receiver. The received pattern is then

$$F(I_0) \sum_n a_n \exp(jn\phi) \cdot \exp - (jn\phi_0), \qquad (20)$$

which is exactly the linear-antenna pattern with amplitude $F(I_0)$. If $F(I) = 1$ for all I (infinite clipper), then a normalized pattern is obtained for any input-signal amplitude.

If two signals are present, then

$$\begin{aligned} I_n &= I_1 \exp(jn\phi_1) + I_2 \exp(jn\phi_2) \\ &= I_1 \exp(jn\phi_1)[1 + a \exp(jn\alpha)], \end{aligned}$$

where I_1 is the larger signal, $a = I_2/I_1$ and $\alpha = \phi_2 - \phi_1$. In this case, the signal no longer has equal amplitude and linearly increasing phase across the aperture, but has amplitude $A_n = [1 + 2a\cos n\alpha + a^2]^{1/2}$ and phase $\theta_n = n\phi_1 + \arctan [a \sin n\alpha / 1 + a \cos n\alpha]$ at the nth element. The received pattern with infinite clipping is now

$$\sum_n a_n \cdot \exp(j\theta_n) \cdot \exp(-jn\phi_0). \qquad (21)$$

Coherent addition will be obtained only when $\theta_n = n\theta$ for some θ. Since $\exp(j\theta_n)$ is periodic with period 2π, it can be expanded in a Fourier series

$$\exp(j\theta_n) = \exp(jn\phi_1) \cdot \sum_{p=-\infty}^{+\infty} C_p(a) \exp(jpn\alpha), \qquad (22)$$

and the received pattern is

$$\sum_{p=-\infty}^{+\infty} C_p(a) \left\{ \sum_n a_n \exp[jn(\phi_1 + p\alpha)] \cdot \exp[-jn\phi_0] \right\}. \qquad (23)$$

The pattern thus consists of the original pattern with "angle harmonics" having phase gradients which are located uniformly about the larger signal. Each "angle harmonic" is the linear pattern. Fig. 18 shows a computed pattern of a linear array with clipping. The true targets are at 0 and 24, while the peaks at -24, ± 48, etc. are spurious.[36]

A plot of the "angle harmonic" amplitudes $C_p(a)$ as functions of a is shown in Fig. 19. The weaker signal is suppressed about 6 db. The "image" at $\phi_1 - \alpha$ has approximately the same amplitude as the true smaller target.

If the signals are below the noise at the clipper, which is usually the case in a large array, then the resultant of signal and noise will be normalized rather than the signal itself. The signal-to-noise ratio in each channel will be reduced by about 1 db. The signals will still add coherently, since phase is preserved by the clipper, but

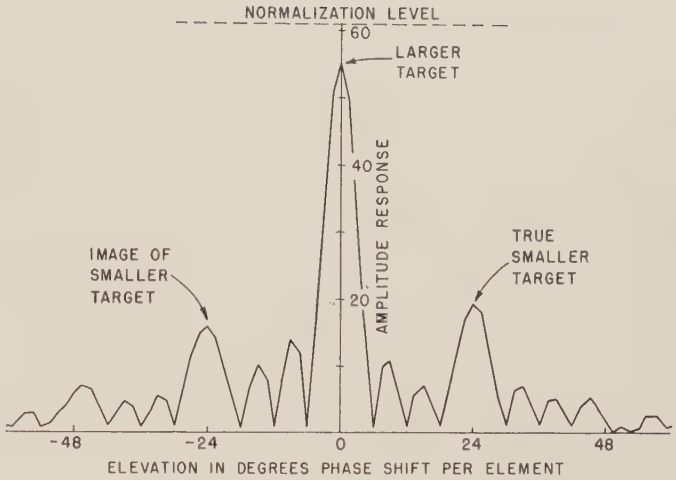


Fig. 18—Computed pattern of 61-element linear array with clipping at each receiver. No taper. Peak output with a single strong signal would appear at "normalization level." Target voltage ratio is 0.6.

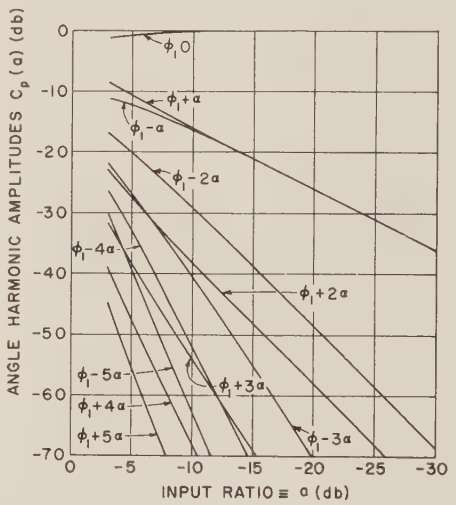


Fig. 19—Angle harmonic amplitudes with two signals and noise neglected. The larger target is at ϕ_1, the true smaller target at $\phi_1 + \alpha$, the first image at $\phi_1 - \alpha$, etc.

the signals will have a random distribution of amplitudes rather than equal amplitude before tapering. The measured pattern of the array will have a random sidelobe level and will be below the normalization level by about the signal-to-noise ratio at the clippers.

Analysis of more complex situations is difficult, but experimental simulation is possible using distinct frequencies into a clipper.[37] This is formally equivalent to the clipped array with multiple signals, with the "pattern" appearing as a frequency spectrum. The computed results of Fig. 19 were verified. Signals below noise at the clippers were simulated with a broad-band noise source and clipper followed by a narrow band-pass filter before spectrum measurement. If two signals are present, both below noise at the clipper, the system becomes more nearly linear in the sense that the relative

[36] The existence of the "angle harmonics" was first pointed out by S. N. van Voorhis. The analysis given here is due to J. C. Nolen.

[37] This method of simulation was suggested by J. C. Nolen and carried out by R. Benson.

amplitudes of the spurious signals decrease. With three signals present, the expected multiple harmonics occur.

When weak signals are present in a clipped array, the array will be approximately linear with a slight gain loss and small spurious responses. As the signals become strong enough to capture the clippers, the weaker signals will be suppressed by about 6 db and large spurious responses will appear, leading to false target generation and angular ambiguity. The clipped array will eliminate an interfering signal in the sidelobe region, but its ability to see main-beam targets in the presence of interference in the sidelobes is somewhat poorer than that of a conventional antenna with an unsaturated receiver.

X. Conclusions

Consideration of the radar requirements listed in Section I has led to the study of large steerable arrays with an amplifier for each element. These devices have the desired properties, but require the use of hundreds or thousands of transmitters and receivers in parallel. Attempts to circumvent the need for large numbers of amplifiers have in general been unsuccessful, although some of the techniques devised for this purpose are useful in special situations.

The purpose of this paper has been to review the principal characteristics of such arrays and the major problems encountered in their design. It is hoped that such a review may stimulate a wider interest in array techniques and lead to better solutions of the inherent problems.

On the theoretical side, further work is needed on synthesis of aperture distributions using unit elements. Nonlinear receiving arrays are known to have interesting properties, and a general theory of gain and resolution for them would be of great interest. Further consideration of the relationships between gain, pattern, receiving aperture and element coupling would be useful.

The future of the array radar depends primarily on component development. The components used in an array radar must meet all of the performance specifications for conventional radar components, and in addition they must be matched over a band of frequencies and must remain stable over long periods of time. A considerable price reduction is also necessary if the large array radar is to be economically feasible.

XI. Acknowledgment

The author wishes to thank his associates at the Bendix Corporation, especially J. C. Nolen and J. H. Best, and members of the engineering staff of Sanders Associates, Inc., for much of the material presented in this paper. He also acknowledges his indebtedness to D. D. King, S. J. Rabinowitz, and S. Falconer for their useful suggestions, and to C. M. Tennant, M. A. Abbott, and C. M. LaPorte for preparing the manuscript for publication.

Signal and Data-Processing Antennas*

G. O. YOUNG†, senior member, ire, and A. KSIENSKI‡, member, ire

Summary—**This paper treats the antenna as an information processing device, and applies the concepts of modern information theory to the design of antennas and to the optimization of their performance. The principal optimization criterion employed is maximization of information or data rate.**

The general procedure is to treat the antenna as a spatial frequency filter which is being optimized subject to a given set of control inputs. Given these specifications, the information rate is maximized with respect to the antenna system parameters subject to the physical constraints of the system.

It is shown that in a general antenna system where noise is introduced in both the object and image space, the optimum antenna aperture distribution is uniform. When the image, or receiver, noise is zero, the useful output information content and rate are independent of the aperture distribution. An equation relating the signal and noise spectra and the aperture distribution is derived which shows the way in which the signal should be coded so as to maximize the information content. Processing is discussed generally, and a specific nonlinear processing scheme is analyzed. The general conclusion is that nonlinear processing degrades the useful information rate when the SNR is low whereas it may improve the rate at high SNRs. Finally, a number of specific military and space applications of information processing antennas are considered.

* Received by the PGMIL, January 28, 1961. The research reported in this paper has been supported by the Electronics Res. Directorate of the AF Cambridge Res. Ctr., ARDC, Bedford, Mass., under Contract No. AF19(604)-3508.

† Aerospace Engrg. Div., Hughes Aircraft Co., Culver City, Calif., and Dept. of Elec. Engrg., University of Southern California, Los Angeles, Calif.

‡ Aerospace Engrg. Div., Hughes Aircraft Co., Culver City, Calif.

I. Introduction

Evaluation of Existing Data-Processing Systems

OVER the past few years the emphasis in antenna design has been shifting from synthesizing patterns to the design of efficient information processing devices. This emphasis has produced numerous schemes which claim to exceed the performance of con-

ventional antennas. Some of these schemes obtain their unconventional performance by nonlinear processing while others make use of increased bandwidth or time modulation. Most nonlinear processors have the advantage over conventional antennas of increased directivity which is obtained at the expense of degradation in gain, multiple-target resolution, or SNR. In general, a trade-off is involved. The SNR is related to the gain and multiple-target resolution capabilities of the antenna. The signal-to-receiver noise ratio is degraded by low gain. The extraction of signal from background noise is deteriorated by poor multiple-target resolution. Typical nonlinear antennas rely on pattern multiplication to remove ambiguities due to spurious beams. As a consequence, the undesired energy is rejected rather than converted into useful energy concentrated in a single beam as in a conventional linear antenna. For a linear antenna, a knowledge of the response to a point target is sufficient to describe the response of any target. This is no longer true for nonlinear antennas where products are formed.[1] The product operation is common to practically all nonlinear antennas. A product results in simple pattern multiplication only in the case of a single-point target. In numerous schemes which have excellent single-target response, multiple-target resolution is poor and the response may indicate targets at wrong positions or none at all.

Several schemes were recently suggested which employ additional parameters to enhance the performance of the antenna. One scheme[2] employs wide-band signals and matched filters to achieve additional directivity, while preserving the linearity of the system. Another scheme[3] modulates the antenna parameters as a function of time to obtain sidelobe suppression or simultaneous scanning.

The fundamental question which arises in connection with all of the above schemes is whether or not the overall performance of the system has improved. In order to answer this question, a basic criterion must be formulated which applies to all systems.

The Information Theory Approach to Antenna Design

The fundamental point of view taken in this paper is that the antenna is an information processing device. The antenna is not treated as a separate entity which must satisfy certain requirements on beamwidth, sidelobe level, etc. but rather as an integral part of a system which attempts to process the information available in an optimum way. Consequently, the concepts of information theory are applicable to the problem of optimization of antenna systems.

One of the principal difficulties in system synthesis and design is the proper specification of an optimization criterion. If the information that the system is intended to yield is properly specified, then surely the "best" optimization criterion is one which maximizes this desired information content (or rate). In the case at hand, the criterion will be that the maximum number of elements in the object be distinguished (*i.e.*, resolved) in the image during a given observation time. Consequently, information rate is the quantity to be maximized in antenna system design.

Surface surveillance radars (or search radars in which the volume to be searched contains a very large number of targets) require higher information rates than search radars (in which the search volume contains few targets) and track radars. It can also be shown that search radars require higher rates than track radars.[4] In tracking radars, only a few bits of information per pulse and/or per lobing cycle are required. The higher the information rate, the greater the required input SNR.

It is difficult to specify properly the minimum distinguishable element. The size of the minimum distinguishable element is a function of the SNR. If there were no noise, infinitesimal elements could be resolved, and the information content would be infinite. Since this ideal situation does not occur in nature, the idea of infinite information content is of academic interest only. A more sophisticated approach is to find the useful information in the image from a knowledge of the probability distributions of the various sources of signal and noise and the effect of the antenna on the average information content, or "entropy," and the rate. The system can then be optimized by maximizing the entropy in the image and the information rate at the antenna terminals with respect to the antenna parameters. Further improvement can be achieved by appropriate processing or coding operations.

Radar Systems Analogies to Classical Communications Systems

It is convenient to apply classical (Shannon) communication theory[5] to antenna design. The classical Shannon communication system is represented in Fig. 1. In this system the nature of the source is assumed to be known completely (that is, its statistics are known). The encoder includes the transmitter, the modulator, and whatever actual encoding scheme is used preceding transmission. The link is the actual transmission channel between transmitter and receiver. The noise is usually assumed to have a constant spectral density over a finite bandwidth, and is purely additive to the signal in the link. The decoder performs the inverse

[1] C. J. Drane, Jr., "Phase-Modulated Antennas," AF Cambridge Res. Ctr. Bedford, Mass., Tech. Rept. No. AFCRC-TR-59-13 (AD215374); April, 1959.

[2] L. R. Dausin, K. E. Niebuhr, and N. J. Nilsson, "The Effects of Wide-Band Signals on Radar Antenna Design," presented at WESCON, San Francisco, Calif.; August, 1959.

[3] H. E. Shanks and R. W. Bickmore, "Four-dimensional electromagnetic radiators," *Canad. J. Phys.* vol. 37, p. 263; March, 1959.

[4] G. O. Young and A. Ksienski, "Signal and Data Processing Antennas," Hughes Aircraft Co., Culver City, Calif., Sci. Rept. No. 8236/1; April, 1961.

[5] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," The University of Illinois Press, Urbana; 1959.

operation to that of the encoder and includes the receiver and whatever demodulating and decoding apparatus is necessary. It is clear that if the antenna system to be designed is to be used in a communication system, the diagram in Fig. 1 is applicable. An analogy between a passive radar system and Fig. 1 is illustrated in Fig. 2. The target in Fig. 2 is analogous to the source in Fig. 1.

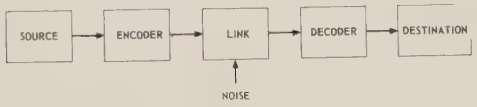


Fig. 1—Communications system.

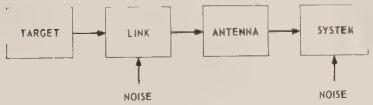


Fig. 2—Passive radar system.

In the passive case, the source is said to be "uncontrolled." In other words, the target may be a planet or heat source which is a true radiator, or it may be a reflector which reradiates incident radiation in such a way that the reflected radiation is completely uncorrelated (*i.e.*, unrelated to) with the incident radiation. Under these circumstances, the radiation from the target cannot be controlled, hence there is no encoder. The link may be considered as the region of space between the target and the antenna. Noise may indeed be added in this region, and may rise from extraneous targets, ground return, background noise, etc. The decoder consists of the antenna and the receiver including computing, tracking, or display apparatus. In general, there will be noise added in the receiver or even in the antenna and RF equipment preceding the receiver as well as in the link. The source itself is usually noisy. This noise exhibits itself in a number of ways. For example, if the source is of finite extent and is scintillating, then the direction of arrival of the wavefront will vary in a random way so that the precise direction of arrival emanating from the center of echo of the target is masked. The received signal may not only scintillate in angle but also in any of the other coordinates (for example, in amplitude). Amplitude scintillation makes the target more difficult to detect in the presence of additive amplitude noise such as is found in the receiver. The target may also scintillate in range if it has a finite depth. Scintillation may also occur in all the derivatives of both angle and range. The target may vary all the way from a perfect reflector (*i.e.*, a controlled source) to a completely uncontrolled source as in the passive case. A perfect reflector simply reradiates the incident radiation with a known change in amplitude and phase and introduces no noise. Such reflections are said to be "specular" and the target is said to be "coherent." If the target is illuminated but the reradiation is not related to the incident radiation, the reflections are said to be "diffuse" and the target is "incoherent." In the most general case, the target produces partially specular and partially diffuse reflection. Such a target is said to be "partially coherent." In order to set up a mathematical model for radar or communications systems, and in an effort to make an analogy to the Shannon communications system, it is necessary to set up a mathematical expression which relates the object or source to the image or antenna output. This is the subject of the next two sections.

II. Spatial Frequency Representation

Let $O(x, y)$ represent field strength in the "object plane," *i.e.*, the plane of the source, or "object." x and y are the coordinates in the object plane. Let $I(x', y')$ represent field strength in the image plane, where x' and y' are the coordinates in the image plane. If the source is a point target, the image is given by the antenna pattern, that is, the "spread function" of the aperture field. When the antenna is scanned, this spread function $h(x, y)$ is convolved with the object field $O(x, y)$ to yield the image field $I(x', y')$,

$$I(x', y') = \int\int_{-\infty}^{\infty} O(x, y)h(x' - x, y' - y)dxdy. \quad (1)$$

Whereas the image plane is real in mirror-type antennas such as parabolic reflectors, the image plane is hypothetical in antennas such as arrays, where the actual antenna output is a vectorial summation of voltages induced on the elements of the array. Nevertheless, the antenna can usually be represented by operations of the type in (1) followed by appropriate linear and/or nonlinear operations on the outputs of the individual elements of the array. If only one dimension is of importance, as in linear arrays, (1) reduces to

$$I(x') = \int_{-\infty}^{\infty} O(x)h(x' - x)dx. \quad (2)$$

Eq. (2) is identical in form to the response of a linear filter to an input time function. $h(x)$ corresponds to the weighting function of the filter, *i.e.*, the response to a unit impulse or delta function. Let

$$\mathcal{O}(u, v) = \mathfrak{F}[O(x, y)] = \int\int_{-\infty}^{\infty} O(x, y)e^{-j(ux+vy)}dxdy \quad (3)$$

where $u = 2\pi f_x$, $v = 2\pi f_y$. f_x and f_y are spatial frequencies. Similarly,

$$\mathcal{I}(u, v) = \mathfrak{F}[I(x', y')] = \int\int_{-\infty}^{\infty} I(x', y')e^{-j(ux'+vy')}dx'dy', \quad (4)$$

and

$$T(u, v) = \mathfrak{F}[h(x, y)] = \int\int_{-\infty}^{\infty} h(x, y)e^{-j(ux+vy)}dxdy. \quad (5)$$

Transforming both sides of (1) yields

$$\mathcal{G}(u, v) = T(u, v)\mathcal{O}(u, v). \tag{6}$$

In the network analogy $T(u, v)$ is now the transfer function of the spatial frequency filter. Note, however, that $T(u, v)$ is equal (except for a constant) to the aperture illumination since it also is related to its far-field pattern by a Fourier transform.[6] The coordinates in the image pattern are related to the angular coordinates of the antenna pattern by

$$x' = \frac{2\pi}{\lambda} \sin\theta' \cos\phi'$$

$$y' = \frac{2\pi}{\lambda} \sin\theta' \sin\phi',$$

where ϕ', θ' and R' are spherical coordinates of the point (x', y') measured from an origin which is the intersection of the antenna axis and the plane of the antenna aperture.

There is one important difference between the single-dimensional and two-dimensional transform relationships. In the single-dimensional case, one may vary the phase along the array in an arbitrary fashion without invalidating the transform relationship between the far-field pattern and the aperture distribution. The phase behavior in the two-dimensional case is restricted to being either constant or varying linearly across the aperture. This, however, does not restrict the spectral behavior or requirements of the far-field pattern. Difficulties may arise in implementing an antenna design which is required to produce a far-field pattern with a given transform. The antenna can still be synthesized by producing a far-field pattern which has the desired transform. The aperture distribution is obtained, however, by solving an integral equation rather than inverting a transform. Certain analogies exist between the temporal spectra of network filters and the spatial spectra of spatial filters. In both cases, the best possible filter characteristic for fidelity of transmission is a flat spectrum. Also, the peak signal-to-average noise ratio is maximized by employing a matched filter. A matched filter has the same attenuation characteristic and the negative phase characteristic of that of the signal. Translating these requirements into the spatial domain requires that the aperture shading be the same as the excitation produced by the source and that the phase distribution be the complement of the arriving phase. In the simple case of a point target, this is equivalent to scanning the antenna in the direction of the arriving wavefront.

In an active coherent radar system, the target is illuminated by an antenna whose far-field pattern is given by $h_1(x)$ and the echo is received by an antenna whose pattern is $h_2(x)$. The image equation is given by

$$I(x') = \int h_1(x' - x)h_2(x' - x)O(x)dx, \tag{7}$$

or if the same antenna is used for transmitting and receiving,

$$I(x') = \int h^2(x' - x)O(x)dx. \tag{8}$$

Thus the spatial frequency characteristic of the antennas in the two-way process is given by

$$\mathcal{G}(u') = \left[\int T(u' - u)T(u)du\right]\mathcal{O}(u'), \tag{9}$$

which is equivalent to passing the signal twice through the same filter. An important difference between temporal and spatial signals is that a time signal has its principal spectral power confined to a limited band, whereas spatial objects are finite and therefore have infinitely wide spectra. In the time domain, the noise spectrum is ordinarily much wider than the signal spectrum so that increasing the filter bandwidth indefinitely eventually worsens the SNR. In contrast, increasing the aperture size always improves the spatial SNR if coherent detection is used.

III. Extension of Information Theory to Space Dimensions and Antennas

The ordinary communications system concerns itself with the properties of time functions and not with the spatial dimensions involved. Antenna system design, however, is concerned with both time and space functions. Hence, the basic theory must be extended to include such functions. Let the angular bounds on the source be $-A \leq x \leq A$, $-B \leq y \leq B$. Let

$$\mathcal{O}_{AB}(u, v) = \int_{-A}^{A} dx \int_{-B}^{B} dy O(x, y)e^{-j(ux+vy)}. \tag{10}$$

Define the spectral density of the object field as

$$G_0(f_x, f_y) = \lim_{\substack{A\to\infty \\ B\to\infty}} \frac{1}{AB} \overline{|\mathcal{O}_{AB}(f_x, f_y)|^2}. \tag{11}$$

The space frequencies f_x and f_y are each measured in cycles per space dimension, for example, cycles per meter. The one-dimensional spectral density $G(f_x)$ is the time and ensemble average of the field strength squared, or average "power," per cycle per meter. In (11), $\mathcal{O}_{AB}$ is the finite spatial Fourier transform of the object field over the limits $\pm A$ and $\pm B$ which enclose the object. The bar denotes an ensemble average. Since the object bounds are usually finite, an approximation to the true density in (11) can be obtained by omitting the limiting

[6] S. Silver, "Microwave Antenna Theory and Design," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 12, p. 174; 1949.

operation. The expression for the spectral density of the image is

$$G_I(f_x,f_y) = \lim_{\substack{\mathfrak{A}\to\infty \\ \mathfrak{B}\to\infty}} \frac{1}{\mathfrak{A}\mathfrak{B}} \overline{|\mathcal{I}_{\mathfrak{A}\mathfrak{B}}(f_x, f_y)|^2}, \tag{12}$$

where $\mathcal{I}_{\mathfrak{A}\mathfrak{B}}$ is the finite transform of $I(x', y')$, and $\pm\mathfrak{A}$ and $\pm\mathfrak{B}$ are the bounds on the image.

If the space frequency bounds on the object are $\pm W_x$ and $\pm W_y$, the object can be completely represented by $16W_xW_yAB$ sampling points in either the space or spatial frequency domains. The sample points in the frequency domain are separated by the reciprocal of the distance between the space bounds in the corresponding dimension, or

$$f_x = \frac{n}{2A}, \qquad f_y = \frac{n}{2B}. \tag{13}$$

Similarly, the image can be represented by $16\mathfrak{A}\mathfrak{B}\mathfrak{W}_x\mathfrak{W}_y$ samples. Note that the number of sampling points representing the object is the same as that representing the image, or

$$16ABW_xW_y = 16\mathfrak{A}\mathfrak{B}\mathfrak{W}_x\mathfrak{W}_y \tag{14}$$

where $\pm\mathfrak{W}_x$ and $\pm\mathfrak{W}_y$ are the bounds on the image spectrum, or aperture. The absolute entropy per object point is by definition

$$H(\mathcal{O}) = -\int p(\mathcal{O}) \ln p(\mathcal{O}) d\mathcal{O}. \tag{15}$$

The successive frequency samples in the object are assumed to be statistically independent and the probability distribution $p(\mathcal{O})$ at each point is assumed to be the same. If the (antenna) filter transfer function is not flat, the entropy per point in the image plane,

$$H(\mathcal{I}_i) = -\int p(\mathcal{I}_i) \ln p(\mathcal{I}_i) d\mathcal{I}_i, \tag{16}$$

is not the same at each sampling point, although the samples are still mutually independent. The average entropy per sampling point is

$$H(\mathcal{I}) = \frac{\sum_{i=1}^{16\mathfrak{A}\mathfrak{B}\mathfrak{W}_x\mathfrak{W}_y} H(\mathcal{I}_i)}{16\mathfrak{A}\mathfrak{B}\mathfrak{W}_x\mathfrak{W}_y}. \tag{17}$$

The useful entropy per image sampling point is by definition

$$H_u = H(\mathcal{I}) - H(\mathcal{I} \mid \mathcal{O}). \tag{18}$$

The conditional entropy $H(\mathcal{I} \mid \mathcal{O})$ is zero if the channel is non-noisy and is equal to $H(\mathcal{I})$ when the system is completely saturated by noise, *i.e.*, when the image and object are completely uncorrelated.

It can be shown[4] that the change in entropy per sample from object to image is

$$H(\mathcal{I}) = H(\mathcal{O}) + \frac{1}{8\mathfrak{W}_x\mathfrak{W}_y} \cdot \int_{-\mathfrak{W}_x}^{\mathfrak{W}_x}\int_{-\mathfrak{W}_y}^{\mathfrak{W}_y} \ln |T(f_x,f_y)|^2 df_x df_y. \tag{19}$$

From (19), the output (image) entropy can be found from the input (object) entropy and the transfer function of the (antenna) spatial frequency filter without explicit knowledge of the probability distribution of the image. This distribution is very difficult to find if the input distribution is non-Gaussian. If the object is noisy, the signal, or desired part of the object, can be called O_S and the noise, or undesired part, O_N. In the general case of additive noise N_2 in the image as well as additive noise N_1 in the object, and assuming N_2 to be statistically independent of the spatial filter output

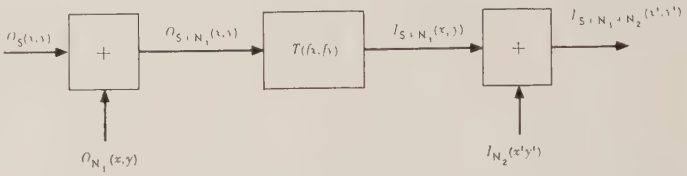


Fig. 3—Functional block diagram of antenna system.

(see Fig. 3) we have,

$$\begin{aligned} I &= I_{S+N_1+N_2} = I_{S+N_1} + I_{N_2} \\ &= \mathfrak{F}^{-1}[\mathcal{O}_{S+N_1}(f_x,f_y)T(f_x,f_y)] + I_{N_2} \\ &= \mathfrak{F}^{-1}[\mathcal{O}_S T] + \mathfrak{F}^{-1}[\mathcal{O}_{N_1}T] + I_{N_2}. \end{aligned} \tag{20}$$

Deriving a general expression similar to (19) is now quite difficult without explicit knowledge of the distributions involved, since in order to make use of the formulas expressing entropy loss after filtering, the entropy for a sum of random variables must be found. The transform of (20) is

$$\mathcal{I} = \mathcal{I}_S + \mathcal{I}_N = \mathcal{I}_S + \mathcal{O}_{N_1}T + \mathcal{I}_{N_2}. \tag{21}$$

It can be shown[4] that

$$H(\mathcal{I}) = H[(\mathcal{O}_S + \mathcal{O}_{N_1})T + \mathcal{I}_{N_2}] \tag{22}$$

and

$$H(\mathcal{I}_n) = H(\mathcal{I} \mid \mathcal{O}) = H(\mathcal{O}_{N_1}T + \mathcal{I}_{N_2}). \tag{23}$$

It can also be shown[4] that

$$G^I_{\mathcal{I}}(f_x,f_y) = G_0(f_x,f_y) \mid T(f_x,f_y)|^2. \tag{24}$$

The total image power is the variance of the image probability distribution, or

$$\sigma^2_{\mathcal{I}} = \int_0^{\mathcal{W}_x} df_x \int_{-\mathcal{W}_y}^{\mathcal{W}_y} df_y G_{\mathcal{I}}(f_x, f_y). \tag{25}$$

If the known input distribution is Gaussian, the entropy for the sum of two variables is readily found, since the output distribution is also Gaussian. Thus, from (21), if the distribution of $\mathcal{O}_s$, $\mathcal{O}_{N_1}$ and $\mathcal{O}_{N_2}$ are all Gaussian and independent, the distributions of $\mathcal{I}$ and $\mathcal{I}_N$ are also Gaussian. If $T(f_x, f_y)$ is not flat, the output spectrum is flat only over the infinitesimal region $\Delta f_x \Delta f_y$. Since the output frequency samples are independent, the total useful entropy can be found by integrating the entropy for each frequency sample over $\pm\mathcal{W}_x \pm \mathcal{W}_y$. Using 21, 24, 25, and the expression for the useful entropy of Gaussian signal and noise,[5] the average entropy per degree of freedom can be shown to be.[4]

$$H_u = \frac{1}{8\mathcal{W}_x\mathcal{W}_y} \int_{-\mathcal{W}_x}^{\mathcal{W}_x} df_x \int_{-\mathcal{W}_y}^{\mathcal{W}_y} df_y \cdot \ln\left(1 + \frac{G_S(f_x, f_y) \mid T(f_x, f_y)\mid^2}{G_{N_1}(f_x, f_y) \mid T(f_x, f_y)\mid^2 + G_{N_2}(f_x, f_y)}\right). \tag{26}$$

A number of interesting and useful conclusions can be drawn from (26). For one thing, the useful statistical information depends only on the magnitude (squared) of the Fourier transforms of the object signal field, the object noise field, the image noise field and the antenna pattern. For example, the phase distribution across the aperture has no effect on the minimum number of distinguishable elements in the image.[7] The phase distribution does affect the antenna pattern, however. A linear phase distribution usually yields that pattern which is most convenient for measurement but does not affect the information content of the echo. The useful entropy is maximized when the magnitude of the integrand in (26) is maximized. Note that

$$\mid T(f_x, f_y)\mid \leq 1 \tag{27}$$

over the range

$$\begin{aligned} -\mathcal{W}_x \leq f_x \leq \mathcal{W}_x \\ -\mathcal{W}_y \leq f_y \leq \mathcal{W}_y. \end{aligned} \tag{28}$$

If the noise and signal spectral densities are fixed, the integrand in (26) is maximized for $T(f_x, f_y) = 1$ over the region defined in (28). Deviations from the equality sign in (27) will be called "aberrations." If there are additional constraints on $T(f_x, f_y)$ so that $\mid T\mid$ cannot be unity over the region in (28), the maximization problem becomes a calculus of variations extremization problem subject to auxiliary conditions. If the various spectral densities are not fixed, further improvement may be attained. Usually, G_{N_1} and G_{N_2} are fixed, but G_S may be able to be controlled. For example, if the desired object, or target, is coherently illuminated, and the target is a specular reflector, by appropriate coding of the transmitted signal the spectrum G_S may be maximized with respect to G_{N_1}. In any case, coherent or even partially coherent target reflections are under the control of the system, hence G_S can in principle be changed in such a way as to maximize the useful received entropy. It is the interaction between the image noise spectrum G_{N_2} and the aberrations of the transfer function that causes the antenna to reduce the useful entropy. If the image noise G_{N_2} is zero, the information per degree of freedom in the image is independent of the aberrations. Indeed, the antenna transfer function does not appear in (26) in this case. In other words, the useful information per sample is completely independent of the antenna pattern when the image is noiseless. Of course, the total entropy, *i.e.*, the total number of distinguishable elements in the image, increases with increase in $\mathcal{W}_x$ and $\mathcal{W}_y$, *i.e.*, with increase in the size of the aperture. In fact, if in addition the signal and noise spectra are flat, (26) reduces to

$$H_u = \frac{1}{2} \ln\left(1 + \frac{\sigma_S^2}{\sigma_{N_1}^2}\right) \tag{29}$$

which is analogous to the Shannon formula for the entropy per degree of freedom for time signals. The general problem of the optimization of (26) is a calculus of variation extremization problem. With $T(f_x, f_y)$ so adjusted as to make $\mid T\mid$ a maximum at every point over the region $-\mathcal{W}_x \leq f_x \leq \mathcal{W}_x$, $-\mathcal{W}_y \leq f_y \leq \mathcal{W}_y$, the integral is maximized for fixed signal and noise spectra. If, however, the signal spectral density $G_S(f)$ is capable of being controlled, further optimization is possible. Assume the commonly occurring situation of a constraint of a given average signal power P_s. This can be written as the integral constraint

$$P_S = \int_0^{\mathcal{W}_x} \int_0^{\mathcal{W}_y} G_S(f_x, f_y) df_x df_y. \tag{30}$$

We wish to maximize (26) after $\mid T\mid$ is made a maximum everywhere subject to constraint (30). The solution to the problem is

$$G_S = -\frac{1}{\lambda} - G_{N_1} - \frac{G_{N_2}}{\mid T\mid^2}. \tag{31}$$

λ is a constant and is adjusted so that (30) is satisfied. Thus, maximum useful entropy is received in the image when the spectra G_S, G_{N_1}, and $G_{N_2}/\mid T\mid^2$ add to give a constant $= -1/\lambda$.

[7] This phase distribution refers to pattern shape and not to scanning which is accounted for in the image equation (1).

IV. Extension to Space and Time Variables

Up to this point, we have simply maximized the total entropy, or information content, in the image, without regard to the time necessary to perform the maximization. The object frame must, however, be scanned in a certain period of time T. It is normally required that the antenna gather as much information as possible in the time T. Thus, the proper optimization criterion for antenna design is maximization of information rate. As the antenna scans the frame, the receiver receives a time-varying message. Let the duration of this message be T and the corresponding (approximate) temporal bandwidth of the message be B_0'. At a particular instant of time, the antenna samples the object. The output voltage across the antenna terminals is proportional to the field strength at the feed if the dimensions of the feed are small compared to the size of the image. In order to reconstruct the object, the antenna scans the object as a function of time. Hence, the image equation is given by

$$I[x'(t), y'(t)] = \int_{-A}^{A}\int_{-B}^{B} O(x, y)h[x'(t) - x, y'(t) - y]dxdy. \quad (32)$$

Thus, although I is a function of t, it is not an explicit function of t as so long as the object remains stationary. The time variation comes about as a result of the scanning. The probability distribution of the image remains stationary, *i.e.*, $p\{I[x'(t), y'(t)]\}$ is not a function of time. Consider a typical variation of image-field strength I vs x' and t as shown in Fig. 4. The samples of I when projected on the Ix' plane are chosen to be $1/(2\mathfrak{W}_x)$ apart, according to the sampling theorem. The corresponding separation of the samples projected on the It plane is $1/(2B_0')$. $V(t)$ is proportional to the projection of $I[x'(t)]$ on the It plane. Since x' and t are functionally related according to the curve in the $x't$ plane,

$$V(t) \sim I[x'(t)] = I(t). \quad (33)$$

The preceding argument can be extended immediately to two space dimensions x and y. In the more general case where the object does move (for example, an aircraft or missile target), $O(x, y)$ is an explicit function of time and hence should be written $O(x, y, t)$. Eq. (32) then becomes

$$I[x'(t), y'(t), t] = \int_{-A}^{A}\int_{-B}^{B} O(x, y, t)h[x'(t) - x, y'(t) - y]dxdy. \quad (34)$$

I also becomes an explicit (as well as an implicit) function of time, and the probability density $pI(x', y', t)$ is nonstationary in time. The image entropy in this case is

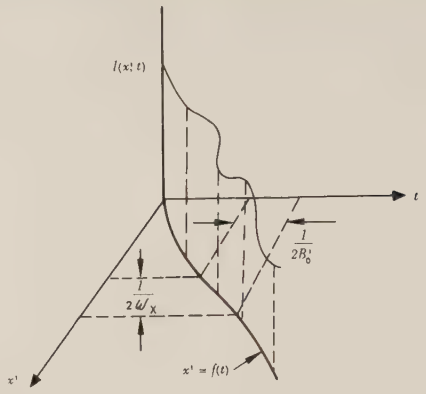


Fig. 4—Representation of image function when object is scanned.

$$H_I(x', y', t) = -\int p[I(x', y', t)] \ln p[I(x', y', t)]dI. \quad (35)$$

If the image is space-stationary, *i.e.*, if the probability distribution at each of the image space-sampling points is the same, H_I is not a function of x' and y'. If, in addition, the image is time-stationary, *i.e.*, if the probability distribution at each of the time-sampling points is the same, H_I is not a function of t. It can be shown[4] that a space average entropy can be defined for space-nonstationary distributions by adding the entropies at each sampling point (if the field strength at each sampling point is mutually independent of that at every other sampling point) and then by dividing by the total number of space-sampling points. In other words, a space average is taken. Likewise, a time average entropy can be defined in the same way by adding the entropies at every time-sampling point and dividing by the number of time-sampling points. By considering a four-dimensional space in which I is plotted vs x', y', and t, a combined space and time average can be taken. Similarly, in the spatial *and* temporal frequency domain, I may be plotted vs f_x, f_y, and f. Since there are $2B_0T$ degrees of freedom in the temporal domain for a message of duration T and bandwidth B_0, the number of degrees of freedom in the spatial and temporal domain are $32\ B_0T\mathfrak{W}_x\mathfrak{W}_y\mathfrak{A}\mathfrak{B}$. Assuming successive frequency samples to be statistically independent and the output or image temporal spectrum to be nonflat, the procedure is similar to that used to derive (26). Note that no additional degrees of freedom occur when the object is not moving, since even if the antenna is scanning, to every t there corresponds a unique value of x', y'. Since information rate is, by definition, the total useful entropy divided by the observation time T, and since time does not appear in the entropy expression when the object does not move (and the image noise is time-invariant), the rate is maximized in this case by letting $T \to 0$. When the object moves, or when the image noise is time-varying, $2B_0T$ additional degrees of freedom appear in the sample space. The total useful entropy is then analogous to (26) except for the additional time variable and degrees of freedom.

Letting f be the temporal frequency variable, the information rate is

$$R = \frac{H_{ut}}{T} = 2\mathcal{A}\mathcal{B}\int_{-B_0}^{B_0} df \int_{-\mathcal{W}_x}^{\mathcal{W}_x} df_x \int_{-\mathcal{W}_y}^{\mathcal{W}_y} df_y \ln\left[1 + \frac{G_s(f_x, f_y f)}{G_{N_1}(f_x, f_y, f) + \frac{G_{N_2}(f_x, f_y, f)}{|T(f_x, f_y, f)|^2}}\right]. \quad (36)$$

The rate in (36) can be maximized in the same way as the entropy in (26).

V. Processing

Since the outputs of the individual elements of an array (each element may itself be an array) are available, it is possible to process the outputs separately before summing. A common class of processing techniques involves successive product operations. These nonlinear operations are usually performed for the purpose of increasing the effective antenna aperture without increasing its physical size. The useful entropy and rate for such processors can be calculated. Consider two arrays with distance "d" between phase centers whose outputs are multiplied as shown in Fig. 5. In the absence of noise, the output voltage for a point target can be shown to be

$$V_f = \frac{\sin 4a_1 x}{4a_1 x} \qquad \left(a_1 = a_2, b = 0, x = \frac{2\pi}{\lambda}\sin\theta\right). \quad (37)$$

Eq. (37) shows the effective aperture dimension to be twice the actual aperture length. For an extended target (or multiple targets), the outputs of the two arrays are, respectively,

$$I_1(x', t) = \int_{-A}^{A} O(x, t) h_1(x' - x) dx \quad (38)$$

and

$$I_2(x', t) = \int_{-A}^{A} O(x, t) h_2(x' - x) dx, \quad (39)$$

where

$$h_1(x) = A_1(x) \sim \frac{\sin a_1 x}{a_1 x} \quad (40)$$

$$h_2(x) = A_2(x)\cos(dx) \sim \cos a_2 x \cos(dx) \quad (41)$$

where

$$d = 2a_1 = 2a_2 = 2a.$$

In this case, the voltage $V_f = V_{f_1} V_{f_2}$, where V_{f_1} is proportional to $I_1(x', t)$ and V_{f_2} is proportional to $I_2(x', t)$. To find the entropy of V_f, the distribution of the product must be evaluated. It can be shown[4] that the distribution of the product of two random variables is

$$p(V_f) = \int p_{V_{f_1}, V_{f_2}}\left(V_{f_1}, \frac{V_f}{V_{f_1}}\right) dV_{f1} \quad (42)$$

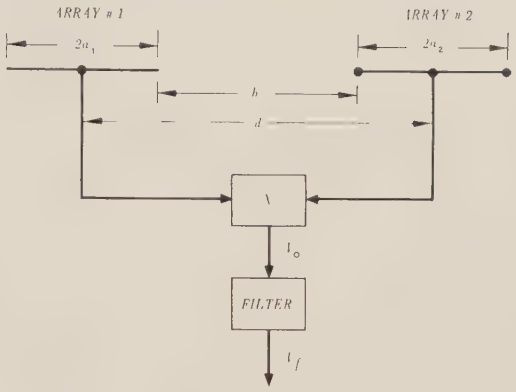


Fig. 5—Simple processing scheme.

where $p_{V_{f_1}, V_{f_2}}$ is the joint distribution of V_{f_1} and V_{f_2}. If the joint distribution of $O(x_1)$ and $O(x_2)$, where

$$\left.\begin{aligned} x_1 &= x(t_1) \\ x_2 &= x(t_2) \end{aligned}\right\} \quad (43)$$

is Gaussian, then $p_{V_{f_1}, V_{f_2}}$ is also Gaussian. The joint distribution of V_{f_1} and V_{f_2} is specified by the means, variances, and correlation coefficient. The means are assumed zero and the variance calculations are straightforward, but the correlation coefficient depends on the correlation of the wavefronts as they simultaneously impinge on the two arrays. Several cases have been calculated[4] for various degrees of coherence. The evaluation of the integral in (42) is very difficult, but has been carried out numerically for several cases.[4] The useful entropy can be calculated for the general situation shown in Fig. 4. Such calculations have been carried out,[4] with the conclusion in each case being that if the SNRs are small, the information content and information rate are degraded by multiplicative antenna processing of the type just considered.

Consider a simple example of independent Gaussian signal and spatial image noise through a flat one-dimensional spatial filter (*i.e.*, a uniformly illuminated antenna aperture). In this case, (36) reduces to

$$R = 4B_0 \mathcal{W}_x \mathcal{A} \ln(1 + \sigma_S^2/\sigma_{N_1}^2) \quad (44)$$

where

$$\sigma_s^2/\sigma_{N_1}^2 = SNR \text{ (in power)}. \quad (45)$$

At SNRs above a given threshold (on the order of unity), the rate R increases more rapidly with increase in bandwidth $\mathcal{W}_x$ than with increase in SNR. At low SNRs, however,

$$R \cong 4\mathcal{W}_x \mathcal{A} B_0 \cdot \frac{\sigma_S^2}{\sigma_{N_1}^2}, \quad (46)$$

so that the rate increases linearly with an increase in either bandwidth or SNR. Since most nonlinear processors gain bandwidth at the expense of SNR, the infor-

mation rate is degraded because of loss in SNR at the same time it is improved because of increase in effective bandwidth. Since at high SNRs, bandwidth increase is more significant than SNR loss, nonlinear processing may provide improvement. At low SNRs, SNR degradation is more effective than bandwidth increase so that nonlinear processing will not improve the information rate, and generally will degrade it. There are many other types of processors, not all of which are nonlinear. An example of a linear processor is the matched-filter technique, wherein a matched filter is used to accept broad-band signals and to discriminate against noise. Linear processors usually discriminate against noise at the expense of increased integration time, or redundancy. Such processors are in general more effective than nonlinear ones at low SNRs, where increase in SNR is as effective in increasing the information rate as the increase in observation time is in decreasing the rate.

VI. Space Coding

One can improve the SNR of time-varying signals by increasing the signal bandwidth (coding) before exposing the signal to noise contamination. When the signal is received, a coherent detection process (decoding) converts the signal to its original bandwidth while discriminating against the noise. The aforementioned tradeoff between SNR and bandwidth assumes that adequate bandwidth is available in the link as well as in the transducer which transfers the signal from the source to the link. Although this assumption is valid for most time signals, it is not generally true for space signals. In the spatial case, the bandwidth available is (defined and in fact) equivalent to aperture size and any manipulation of the far field (which is the equivalent of varying the time function) cannot increase the spatial bandwidth of the antenna which is continually used to its full extent.

Redundant coding is useful in the reduction of errors. One type of error is due to angular scintillation which causes the apparent location of a target to be different from its true location. This kind of error is especially important in systems using simultaneous scanning, where the whole observation range is illuminated simultaneously. The echo may then be received on the wrong beam. This difficulty may be remedied by transmitting the various beams at slightly different frequencies.[3] If a stacked beam system is used for receiving, the signal frequency at each pair of terminals is compared with the frequency which should appear at those terminals according to the code. This error detecting and correcting code substantially reduces scintillation errors.

VII. Applications of the Theory

Substantial improvement over conventional techniques in planetary surface surveillance and mapping can be achieved by maximizing information rate according to the procedures developed. Since in this case the object dimensions are large compared to the beamwidth, the antenna can be designed to yield the maximum magnitude of the spatial frequency transfer function, and the scanning can be programmed in such a way that (31) is satisfied. By varying the intensity of the beam as it scans the spectrum, $G_S(f_x, f_y)$ can be appropriately coded such that,

$$G_S(f_x, f_y) + G_{N_1}(f_x, f_y) + \frac{G_{N_2}(f_x, f_y)}{|\, T(f_x, f_y)\,|^2} = \text{constant}. \quad (47)$$

As an example, consider an extended object whose surface is specular (a smooth surface) over a portion of the area and diffuse (rough surface) over another. The optimum scanning program to yield the clearest picture of the object for a given observation time would spend more time illuminating the specular area than it would the diffuse area, thus increasing the SNR of the integrated return signal. The computer which will perform this function must be quite sophisticated so as to adjust the program according to the previously received information. In other words, the system should be self adaptive. Improvement according to (47) can be achieved if the return from the object is at least partially coherent.

Space coding such as described in this paper can be advantageously used in a number of technical military applications. For example, a search-while-track radar can employ simultaneous scanning and, at the same time, discriminate against unwanted targets and angular scintillation by frequency coding the beams. Undesired targets, decoys, and other active countermeasures can be discriminated against once the desired target has been selected and appropriately coded.

Improvement in the resolution of receiving antennas used in tracking small targets such as satellites or I.C.B.M.'s can be achieved by appropriate processing. In satellite communication systems, directivity is of utmost importance. Extreme directivity can be obtained at the expense of increased error rate in communication. By appropriately coding the source, however, high SNRs can be insured so that the increase in error rate resulting from improved directivity is still within the acceptable tolerance. Essentially infinite range communication systems can be devised by employing satellite repeaters whose separation is determined by the error rate tolerance.

Improved directivity will improve the radar and communication range of limited diameter and limited weight vehicles such as satellites by means of lightweight, compact processing devices instead of by brute force methods such as increasing transmitter power and weight.

The principal purpose of this paper has been to determine methods for realizing the optimum information processing antenna system yielding the maximum information rate consistent with the system's physical constraints. The second purpose was to indicate the use of these methods in existing and proposed military and space vehicles.

Signal Processing Techniques for Surveillance Radar Sets*

C. A. FOWLER†, MEMBER, IRE, A. P. UZZO, JR.†, MEMBER, IRE, AND A. E. RUVIN†, MEMBER, IRE

Summary—**One of the major recent advances in radar technology has been in the processing of the received signals. Several techniques have been developed to enhance the desired signals (aircraft) relative to ground clutter, sea clutter, rain, interference from other radars, and active countermeasures.**

The following techniques are discussed: MTI, Sweep Integration, and Blanking and Switching. A signal processing system utilizing these techniques is described.

INTRODUCTION

OVER the past decade, one of the major advances in radar technology has been in the processing of radar signals. Technique developments have been principally associated with the relief of effects due to ECM and ground clutter and improvement in target detection capability. Additional effort has also gone into the development of automatic and semi-automatic devices that select the best signal processing techniques and output signals for use in particular areas and in specific environments with emphasis on the multiple processing of the radar video.

Signal processing techniques can be divided into two broad classes—techniques that are essentially linear and attempt to separate the desired signals from interference or noise on the basis of some selective property of the system, and techniques that are based on blanking of the radar output in the presence of undesired signals. Both approaches have areas of utility, but it is preferable to use enhancement of the desired signals prior to the use of blanking.

Radar returns of interest can be characterized as having a known frequency, spectrum, pulse length and repetition period.

MTI

MTI (Moving Target Indicator) is a radar technique used to separate moving targets from fixed targets. The signal received from a moving target is essentially a line spectrum shifted from that of fixed targets by the Doppler effect. In an MTI, the composite signal of fixed clutter and moving targets is processed to filter out the fixed target spectral components and to pass the others.

The MTI, developed during World War II for surveillance radar,[1] was a coherent, uniform PRF system utilizing a simple single canceler as a comb filter (Fig. 1). The velocity (frequency) response of this comb filter is shown in Fig. 2. This canceler has nulls (blind speeds) at multiples of the radar PRF (fr), and thus completely cancels an ideal fixed target with a true line spectrum. However, the scanning of the radar antenna causes the spectrum lines to broaden as shown in Fig. 2. The spectral width is proportional to the tip speed of the radar antenna.

A more ideal response shape would be less sensitive near the nulls and have a more uniform response away from them. A response of this type can be obtained by the use of a velocity-response-shaping-canceler,[2] which consists of two single cancelers in cascade with feedback, as shown in Fig. 3. The shape of the response is controlled by the feedback factors. Fig. 4 shows a velocity

Fig. 1.

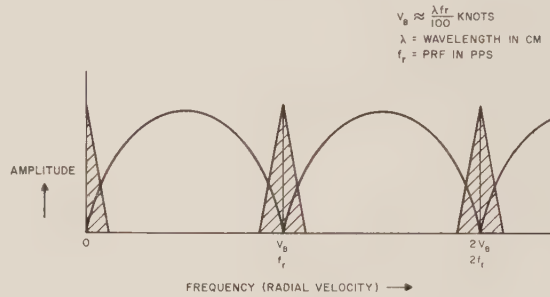


Fig. 2.

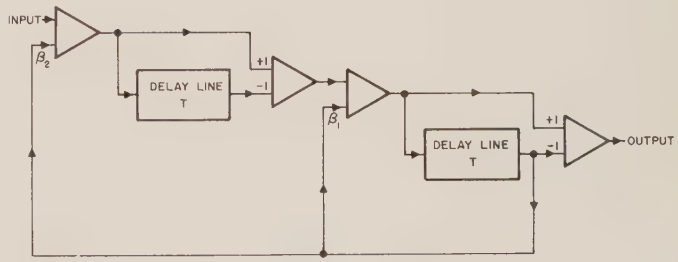


Fig. 3.

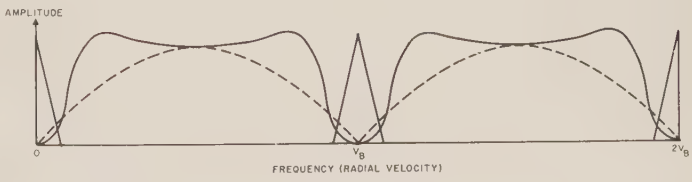


Fig. 4.

* Received by the PGMIL, January 18, 1961.

† Airborne Instruments Lab., Div. Cutler-Hammer, Inc., Melville, N. Y.

[1] L. N. Ridenour, "Radar Systems Engineering," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1, ch. 16; 1947.

[2] W. D. White and A. E. Ruvin, "Recent advances in the synthesis of comb filters," 1957 IRE NATIONAL CONVENTION RECORD, pt. 2, pp. 186–200.

response typically available with this approach. (The response of the single canceler is also shown for reference.)

With this technique the scanning limitation can be readily removed for nearly all surveillance radars. The exact response selected depends upon the scanning clutter and the performance desired. Fig. 5 shows four typical responses each adjusted to give 30-db cancellation of fixed targets for the designated number of PBW (pulses per beamwidth). In many radars the parameters are such that only a modest amount of shaping is needed to reduce the scanning fluctuations. It may still be desirable, however, to have the more severe shapes available to reduce clutter, having additional fluctuation such as rain return or sea clutter. For example, the rain return on a 400 PRF, *L*-band radar is reduced 30 db with the 5.8 PBW shape compared with 15 db obtained from a single canceler.

This example assumed that the fluctuation spectrum of the rain is centered about the blind speeds. Therefore, to be valid, the group velocity of the rain due to wind must be removed. This can be done by the use of noncoherent MTI. This type of MTI uses the clutter (or rain) as the reference signal for moving target detection. Therefore, in the case of a target in rain clutter, it is the radial component of the relative motion of the target to the rain that produces the Doppler beat. (Noncoherent MTI does not discriminate between group velocity of rain and sea clutter and platform motion, and therefore can be readily used on moving platforms such as aircraft and ships to provide MTI capability against clutter.)

The combination of noncoherent MTI with velocity response shaping is therefore a powerful tool in dealing with targets such as rain and sea clutter.

Since noncoherent MTI utilizes the beat between target and clutter, it is "blind" to targets not over clutter and loses sensitivity on targets over weak clutter. It is therefore desirable to use it only in those regions where it is necessary (*e.g.*, rain clutter areas) and to use either coherent MTI or normal video everywhere else in the coverage area. Means for accomplishing this are discussed later.

When severe response shapes are used to suppress extraneous targets, the blind speed zones are widened. This can be alleviated by using a nonuniform interpulse interval—a staggered PRF. Staggered PRF is accomplished by delaying every other transmitted pulse by an amount, ΔT. The ratio of the two unequal interpulse intervals is called the stagger ratio. Upon reception, the signals are destaggered using the same ΔT delay element that generated the original stagger. Fig. 6 shows the timing diagram of a stagger/destagger system.

Fig. 7 shows the over-all velocity response of a 5/7 stagger combined with a velocity-shaping canceler. The gross effect of the stagger is to multiply the first blind speed by the factor $T/\Delta T$. However, the larger this factor is, the deeper the holes in the response curve.

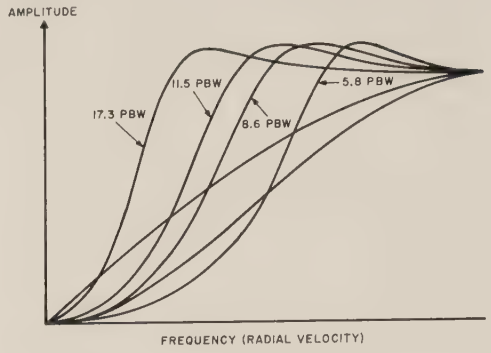


Fig. 5.

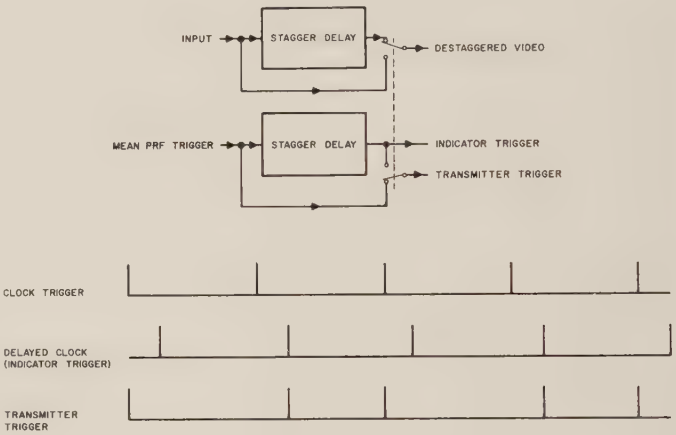


Fig. 6.

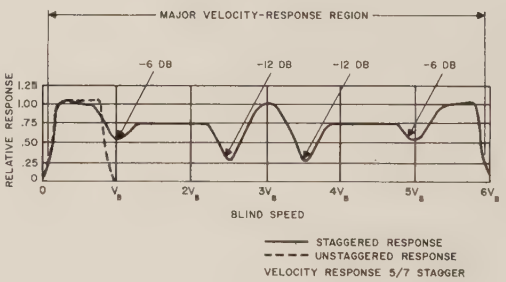


Fig. 7.

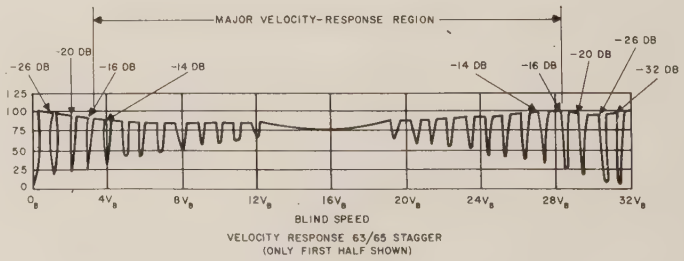


Fig. 8.

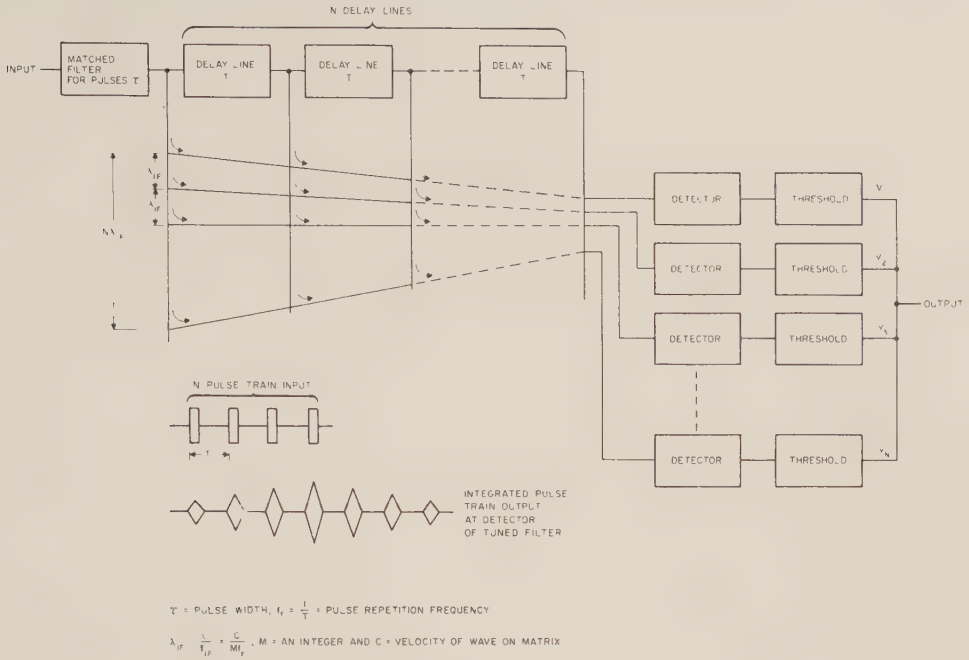


Fig. 9.

This is evident in Fig. 8, which shows the response obtained with a 63/65 stagger ratio. The higher stagger ratios are useful in optimizing the response in a particular velocity region.

Integrators

The threshold detection for a group of weak pulses in a background of random noise can be greatly enhanced by integration.[2] If the integration of the pulses takes place prior to envelope detection, the process is called coherent integration. When the phasing of N successive pulses has been corrected for target motion, the enhancement of the signal power will be N times the noise power. (The signal voltages add linearly while the noise voltages add in an rms manner.) If the integration takes place after detection, the process is called noncoherent integration.

A hypothetical example of a coherent integration bank of filters for the N pulse train is shown in Fig. 9. It is assumed that the individual pulses are optimally filtered by means of a matched IF filter. The concept of N equal pulses from a point target implies a rectangular beam shape. A more realistic example is a uniformly illuminated rectangular aperture resulting in a sin X/X one-way beam pattern or a $\sin^2 X/X^2$ round trip scanning modulation. The signal voltage spectral shape of the line is triangular, and it is not physically realizable for an integrator to have a finite pass band required for a matched filter response. In practice, an optimum 3-db bandwidth for an integrator is approximately fr/2(PBW).

Sweep integrators have been designed around a single period delay line with a positive feedback factor K slightly less than unity enhancing the harmonics of the pulse repetition frequency. Fig. 10 shows the basic block diagram and frequency response of such an integrator. The 3-db bandwidth of such an integrator is approximately $(1-K)\mathrm{fr}/\pi$. The response of such an integrator for an infinite train of optimum Doppler pulses is $1/1-K$, and the noise response is $1/\sqrt{1-K^2}$. The response of the integrator to a finite train of equal level pulses is $1-K^N/1-K$. The signal enhancement of an N pulse train is

$$(1 - K^N)\left(\frac{1 + K}{1 - K}\right)^{1/2}.$$

The feedback sweep integrator exponentially weights the pulse in the train with maximum weighting of the latest pulse. A bank of delay line sweep integrators is required for coherent integration or else the single delay line integrator may be multiplexed either in time or in

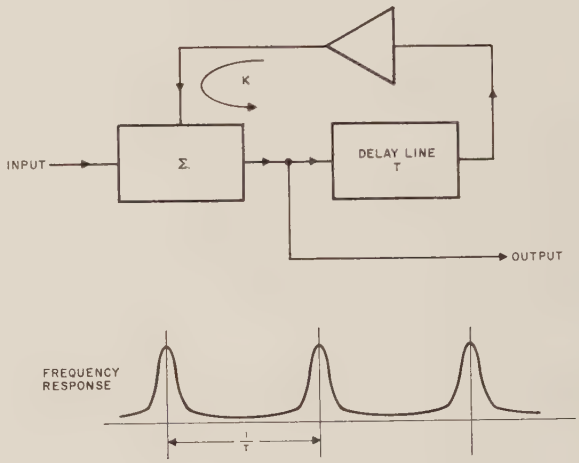


Fig. 10.

frequency if sufficient bandwidth is available.

If Doppler resolution is not required, noncoherent (post-detection) integration may be used. The losses compared with coherent integration are not excessive for scanning radars (less than 100 pulses per beamwidth).[3] Post-detection integration losses of 1.2 db, 3 db, 4.1 db, 8 db and 12.3 db exist for 10, 50, 100, 1000, and 10,000 pulses respectively for a square law detection probability of 0.99 and a false alarm probability of 10^{-10}. A signal-to-noise ratio at the detector input of unity (0 db) results in a signal-to-noise ratio after detection of −4 db. The post-detection integration of 100 pulses results in a 20-db enhancement giving the +16-db signal-to-noise ratio required for 0.99 detection probability. Detector suppression results in a square law relation between output and input signal-to-noise ratios for weak signal-to-noise levels.

A mixture of coherent and noncoherent integration can be employed to reduce the total number of integrators required. For example a bank of 10 coherent integrators followed by a bank of 10 noncoherent integrators (total of 20) is nearly as effective from a detection standpoint as 100 coherent integrators. In addition, the target acceleration in some cases results in a Doppler which precludes the use of the very narrow filters required for fully coherent integration. The coherent integration must be adequate to enhance the signal-to-noise ratio to a level of at least 0 db at the detector.

Sweep integrators are useful for pulse and low duty interference reduction. The signal dynamic range into the integrator is restricted by means of a limiter so that the response to random pulses in the integration time can be restricted to a level below the detection threshold.

The use of an integrator ahead of a PPI does not result in an increase in detection range of a normal search radar because the PPI phosphor is itself a good integrator for this case. An integrator ahead of a PPI, however, results in an improved contrast that is more pleasing to the operator. A black-faced indicator is required in order to avoid any integrated noise painting. Integrated noise, if presented, has the blip dimensions similar to targets. The use of a black-faced tube also allows the easy recognition of target trails.

The stability problems associated with high feedback integrators are serious. An interesting technique (devised by the Philco Corporation) for achieving stability makes use of a double loop. One loop is oscillating (unity feedback) and the other has a small negative feedback factor (β). The result is a high positive feedback factor, $K=1-\beta$. The system shown in Fig. 11 uses frequency modulation of a deviable oscillator. A stable buildup of $1/1-K=1/\beta$ appears at the integrated video output for an infinite unit pulse train.

[3] J. I. Marcum and P. Swerling, "Study of target detection by pulsed radar," IRE TRANS. ON INFORMATION THEORY, vol. IT-6, pp. 59–144; April, 1960.

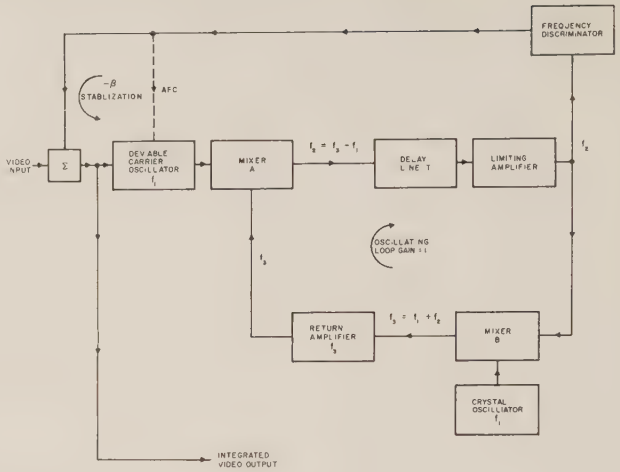


Fig. 11.

Blanking and Switching Techniques

There is a class of video processing devices that can be characterized by the common property of desensitization of the radar output or selection of an alternate signal source whenever activated by an anomaly in the radar video signal: one such device, the pulsewidth discriminator, serves a variety of useful purposes as part of a video processing system. Fig. 12 shows a block diagram of a typical pulsewidth discriminator. The video input is delayed by an amount equal to the desired pulsewidth criterion; that is, by some definition (usually about two pulsewidths) that is chosen as the basis for deciding between desired radar returns and interference or ground clutter, etc. The delayed and undelayed signals are then compared in a coincidence gate that produces an output whenever the input signal exceeds the desired criterion. The output signal from the coincidence gate is equal to the input signal width less a width equal to the criteria and this is added on by the pulse stretcher.

The output signal from the pulsewidth discriminator is usually applied to a video switch. The pulsewidth discriminator can be used as a simple clutter suppressor, blanking out the areas where clutter is sensed, as defined as being targets of excessive pulsewidth. If MTI is substituted in the areas where clutter is sensed, a clutter-gated MTI system is obtained wherein MTI is presented only in the regions of clutter and normal video outside these areas. Since MTI is more susceptible to certain forms of ECM and interference, the pulsewidth discriminator can be used to advantage as a clutter gate to confine MTI operation to the minimum required area. Furthermore, when noncoherent MTI (which is blind in areas of no clutter) is used, clutter gating is a necessity.

A "capture gate" may be used to suppress isolated fixed point targets, which are not gated by simple pulsewidth discriminator. If a PPI picture is examined, it will be found that, almost invariably, heavy clutter areas are followed by several isolated fixed point targets. The capture gate (shown in Fig. 12 within the

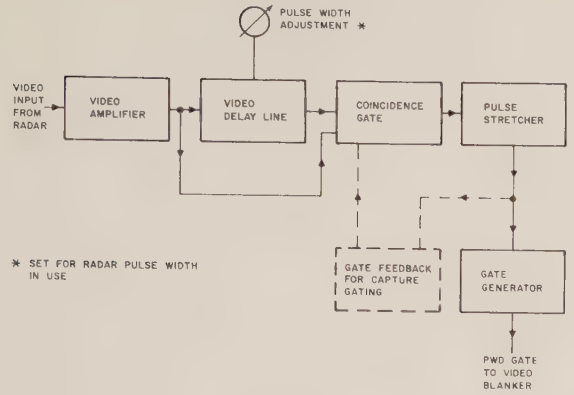


Fig. 12.

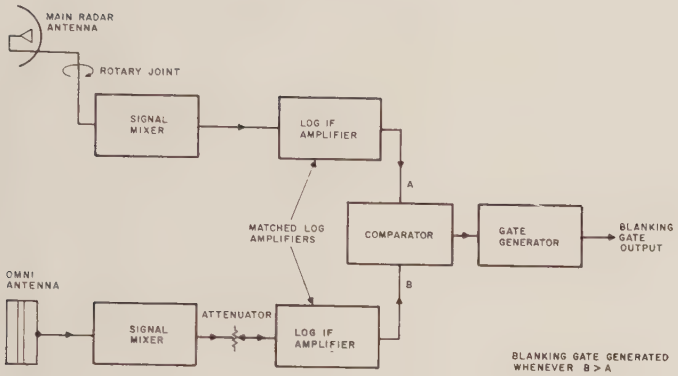


Fig. 13.

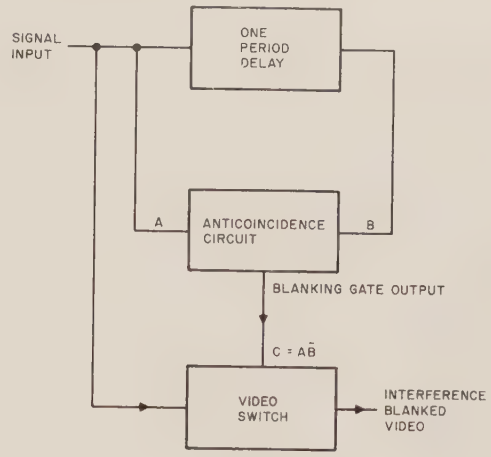


Fig. 14.

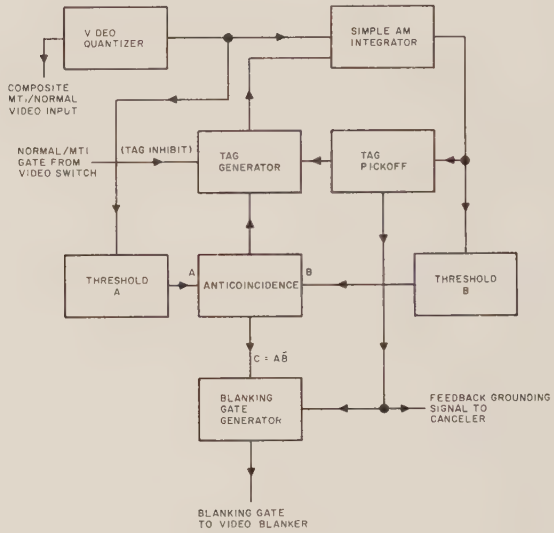


Fig. 15.

dotted lines) basically keeps one line of the coincidence gate activated for a presettable length of time subsequent to detection of a clutter pulse. Thus, if a pulse of any width (for example, an isolated point of clutter) should appear within this interval, it will extend the output gate to include (or "capture") the point target.

Another technique, sidelobe blanking, is analogous to pulsewidth discrimination except that it depends on direction of signal arrival rather than a peculiarity of the signals. The sidelobe blanker prevents the entry of interfering signals via the sidelobes of the radar. These signals arise from adjacent radars, or from intentional attempts to put confusing signals into the radar receiver when the antenna is pointing in the other directions. Fig. 13 shows a block diagram of a typical sidelobe blanker. An omnidirectional antenna and receiver are adjusted so that the gain at all azimuths is slightly in excess of the gain in the main antenna at the peak of the strongest sidelobe. The output of the main radar and omnichannel are compared, and the radar output is blanked whenever the omnireceiver output exceeds the main radar output indicating that the signal is coming in via a sidelobe rather than through the main beam.

Another type of blanking device is the PISAB (pulse interference separator and blanker).[4] The PISAB checks for synchronism on a pulse-to-pulse basis and rejects all signal pulses that do not show the requisite synchronism. Basically, the circuits include a one radar sweep period delay using either ultrasonic or magnetostrictive delay lines or storage tubes. The output of the delay line is compared with the input and the radar video is blanked whenever there is no signal at both the input and output of the delay line, as would be the case for asynchronous signals. Fig. 14 is a block diagram of a simple PISAB. Basic problems that arise are: 1) this simple unit may blank excessively when there are fades, etc., within the pulse train, and 2) it will not work with MTI because an MTI subjected to a single input pulse produces a pulse train of two to ten synchronous pulses depending upon the complexity of the canceler. Fig. 15 shows a block diagram of PISAB which works on both normal and MTI and overcomes most of the other objections to the simple system. The system uses the anticoincidence principle but is supplemented by an inhibiting function which prohibits blanking more than the required number of times (once for normal video

[4] H. Kurland, "Interference blanker for normal and MTI (moving target indicator) video," *Proc. National Electronics Conf.*, Chicago, Ill., October 7–9, 1957.

and three times for a double canceler) at any interfering interpulse azimuth and range. The input video is fed to a video sweep integrator and to threshold A, the output of which feeds the anticoincidence circuit. The other input to the anticoincidence circuit is the output of threshold B. The input to threshold B is the integrated video from the video integrator. The anticoincidence circuit determines what information is to be treated as interference. The output from the anticoincidence circuit then blanks the interference for the first period and activates the tag generator. The function of the tag generator is to blank the interference interval during the required number of succeeding periods, and to provide the means for inhibiting blanking in the following periods.

System Considerations

When attempting to combine the various signal processing devices into a working system, some combinations are obvious and common such as the addition of MTI to a radar. Some radar systems employ noncoherent MTI as a technique against rain and seaclutter, either with or without the addition of coherent MTI for ground clutter. In addition, the pulsewidth discriminator or derivatives of it can be employed for clutter gating, blanking of targets of excessive pulsewidth on both normal and MTI, and elimination of transients. Sidelobe blanking can be included in the system to prevent entry into the radar of undesired signals such as spoofs, repeaters and pulses from adjacent radars.

The filter buildup and decay of a velocity response shaping MTI canceler causes a shift in the azimuthal position of a target. A video integrator can be used in series with the normal line to shift the normal video by a like amount since a video integrator also causes an azimuthal shift. Unfortunately, the feedback factor for azimuthal realignment is not normally high enough to permit effective interference suppression on normal. Furthermore, a video integrator is less effective on MTI. Therefore, a PISAB is usually required to suppress random pulses.

Fig. 16 shows an over-all block diagram of a signal processing system for performance against rain, sea, and ground clutter, some forms of ECM, and pulse interference.

A separate MTI system into which coherent and/or noncoherent video can be processed (but not concurrently) provides an MTI capability against ground

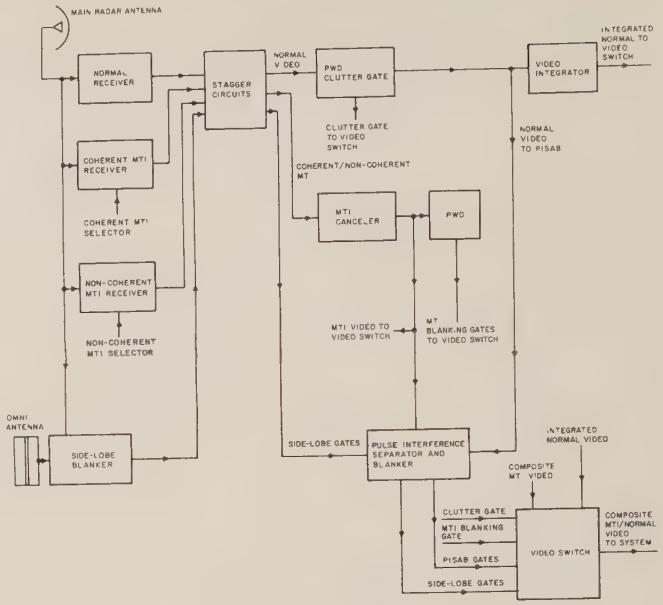


Fig. 16.

clutter (coherent MTI) and rain and/or sea clutter (noncoherent MTI). Coherent MTI video is chosen on a range gated basis (it does not have the problem of blindness in the clear). Beyond this range, the normal video is multiplexed with the noncoherent MTI by means of a pulsewidth clutter gate. Pulsewidth discriminators are associated with the MTI and normal lines to eliminate interference and other residues that fall outside the general pulsewidth of the radar. The normal video line includes a video integrator in order to permit realignment of the normal video with the MTI video. This integration is done subsequent to pulse interference blanking since the integrator will generate a pulse train from a single pulse and prevent effective operation of the pulse interference blanker. The MTI/normal video can be applied to a pulse interference blanker for the generation of the appropriate MTI/normal blanking signals. The blanking gates generated in the various video processing subunits are applied to a common video switch unit to provide a composite blanked video output.

Signal processing systems such as described above are currently being built by radar manufacturers for a variety of radars and in conjunction with some classified techniques can provide an adaptable system that can, to a great extent, reduce or suppress undesired interference.

Principles of Pulse Compression*

H. O. RAMP†, MEMBER, IRE, AND E. R. WINGROVE†, MEMBER, IRE

Summary—For good radar system performance, a transmitted waveform is desired that has 1) wide bandwidth for high range resolution and 2) long duration for high velocity resolution and high transmitted energy. In a pulse-compression system, a long pulse of duration T and bandwidth F (product of T and F greater than one) is transmitted. Received echoes are processed to obtain short pulses of duration $1/F$. Compression ratio (the ratio of long-pulse duration to short-pulse duration) is thus the product $T \cdot F$, and is a measure of the combined range and velocity resolution.

Pulse compression occurs if a waveform with a nonlinear phase spectrum is passed through a filter "phase matched" to the waveform. Phase matched means that the nonlinear part of the network phase response is the negative of the nonlinear part of the waveform phase spectrum.

The waveform with rectangular amplitude spectrum and parabolic phase spectrum is ideally suited for pulse compression. Bandwidth and duration can be independently specified, duration-bandwidth products of 100 or more are presently feasible, and the waveform remains phase matched to one filter over a wide range of Doppler frequency shifts.

Radar Performance and the Signal Waveform

MODERN radar systems must rapidly gather reliable position and velocity data on a large number of targets which can be at extreme ranges. As this task becomes increasingly more difficult, ways to improve radar performance must continually be sought.

Some improvements are still being made in the components which affect radar performance; for example, receivers with lower noise figures, transmitters with higher power output, and antennas with more gain. Further improvement is obtained by means of elaborate signal processing schemes, some of which are optimum filtering, pulse-to-pulse integration, and MTI. However, until recently, very little effort has been expended to improve a factor which constitutes a fundamental limitation to radar performance; namely, the *transmitted waveform*.

The ways in which the transmitted signal limits the radar performance have been determined and are fairly well known.[1] Three parameters are of prime importance. These are the *bandwidth*, the *time duration*, and the *total energy* of the transmitted signal. Specifically:

1) Range Resolution is determined by the frequency structure or spectrum of the signal. For a given spectral shape, range resolution is proportional to $1/F$ seconds where the *bandwidth* of the signal is F cycles. For a given bandwidth, sharpest range resolution is obtained with a signal which has a rectangular amplitude spectrum.
2) Velocity (the radial component of target velocity is meant) resolution is determined by the time structure or envelope of the signal. For a given envelope shape, velocity resolution is proportional to $1/T$ cycles where the *time duration* of the signal is T seconds. For a given time duration, finest velocity resolution is obtained with a signal which has a rectangular envelope.
3) Target detectability is determined by the ratio of received signal *energy* to receiver noise power. For a given antenna gain, receiver noise figure, target size, and signal processing scheme, signal detectability and thus maximum range can be improved only by increasing the transmitted energy.

The most commonly used radar-signal waveform is the signal that results when an oscillator is turned on, then off again. Idealized, this "simple radar pulse" is a number of cycles cut out of a continuous sine wave. If this waveform is viewed in light of the preceding relationships, it is seen to be an impasse. For one thing, its bandwidth and time duration are reciprocally related. Furthermore, if the envelope is made rectangular, the spectrum has a sin x/x-shape; if the spectrum is made rectangular, the pulse has a sin x/x-envelope. With the simple radar pulse, an increase in range resolution can be obtained only by accepting a loss in velocity resolution, and vice versa. Secondly, since the energy of the pulse is the product of the power and the duration, when the transmitter is operated at its peak-power limit (often the case), the energy can be increased only by extending the duration, thus decreasing the bandwidth. Increased maximum range can be exchanged for a loss in range resolution, or, by sacrificing maximum range, range resolution can be increased. With the simple radar pulse, radar performance is a compromise between velocity resolution and maximum range on one hand and range resolution on the other.

A waveform is desired that has both wide bandwidth and long duration, preferably one with a rectangular spectrum as well as a rectangular envelope. The product of duration and bandwidth, or $T \cdot F$, can be defined as a figure of merit or measure of the performance of a particular waveform. However, while a large $T \cdot F$ is necessary for an improvement in performance, it is not sufficient to guarantee it.

* Received by the PGMIL, March 16, 1961. The work on which this paper is based was supported in part by the Air Res. and Dev. Command and the Rome Air Dev. Center.

† Electronics Lab., General Electric Co., Syracuse, N. Y.

[1] P. M. Woodward, "Probability and Information Theory," Pergamon Press, London, Eng; 1957.

Ideally, one would like to design the time waveform and frequency spectrum independently. Unfortunately, when the time function is chosen, the frequency spectrum is of course also determined. Thus, the waveform cannot be designed in a straightforward procedure.

In addition, theoretical considerations must be tempered by practical necessities. Although a particular waveform may have desirable properties, it does not have much value if the equipment required to implement the system becomes absurdly complicated.

Therefore, an improved signal waveform is to be found more by intuitive reasoning based on known facts than by design. There are *two* problems to be solved. First, a waveform capable of providing the required range and velocity resolution must be found. Then, a feasible means must be devised to generate the signal and process it in a manner that extracts all the information it contains.

Nomenclature

Before continuing with the discussion of the problem of waveform selection, the reader should refer to Tables I and II. Here a number of general terms concerning waveforms and networks are defined and symbols for them assigned. These terms prove to be very useful in facilitating the explanations which follow.

No claim is implied that all of the terms listed in the tables are new. The tables are included for completeness and to provide precise meanings for terms which might otherwise be ambiguous. Most of the terms and their

TABLE I

Properties of a Band-Pass Type Waveform

Time Domain		Frequency Domain
Waveform	← →	Spectrum
$a(t-\sigma_0)e^{j[\omega_0\cdot(t-\sigma_0)+\phi(t-\sigma_0)+\phi_0]}$	Fourier ← Transform → Pairs	$A(\omega-\omega_0)e^{j[\sigma_0\cdot(\omega-\omega_0)+\Phi(\omega-\omega_0)+\phi_0]}$
$a(t-\sigma_0)$ = Envelope		$A(\omega-\omega_0)$ = Amplitude Spectrum
t = Time		ω = Frequency
σ_0 = Time Shift of the Waveform		ω_0 = Carrier Frequency
$\omega_0\cdot(t-\sigma_0)+\phi(t-\sigma_0)+\phi_0$ = Instantaneous Phase		$\sigma_0\cdot(\omega-\omega_0)+\Phi(\omega-\omega_0)+\phi_0$ = Phase Spectrum
ϕ_0 = Fixed Carrier Phase		ϕ_0 = Uniform Phase
$\phi(t-\sigma_0)$ = Phase Modulation		$\Phi(\omega-\omega_0)$ = Phase Dispersion
$\frac{d\phi}{dt}+\omega_0$ = Instantaneous Frequency		$-\frac{d\Phi}{d\omega}-\sigma_0=\sigma$ = Phase Slope
ω_0 = Carrier Frequency		$-\sigma_0$ = Uniform Phase Slope
$\frac{d\phi}{dt}$ = Frequency Modulation		$-\frac{d\Phi}{d\omega}$ = Phase-Slope Dispersion

TABLE II

Properties of a Band-Pass Network

Frequency Domain		Time Domain
Frequency Response	← →	Impulse Response
$G(\omega-\omega_c)e^{j[D_0\cdot(\omega-\omega_c)+\Theta(\omega-\omega_c)+\Theta_0]}$	Fourier ← Transform → Pairs	$g(t-D_0)e^{j[\omega_c\cdot(t-D_0)+\theta(t-D_0)+\Theta_0]}$
$G(\omega-\omega_c)$ = Amplitude Response		$g(t-D_0)$ = Envelope of the Impulse Response
ω = Frequency		t = Time
ω_c = Center Frequency		D_0 = Time Shift of the Impulse Response
$D_0\cdot(\omega-\omega_c)+\Theta(\omega-\omega_c)+\Theta_0$ = Phase Response		$\omega_c\cdot(t-D_0)+\theta(t-D_0)+\Theta_0$ = Instantaneous Phase
Θ_0 = Uniform Phase Shift		Θ_0 = Constant Phase of the Impulse Response
$\Theta(\omega-\omega_c)$ = Phase Dispersion		$\theta(t-D_0)$ = Phase Modulation of the Impulse Response
$-\frac{d\Theta}{d\omega}-D_0=D$ = Delay		$\frac{d\theta}{dt}+\omega_c$ = Instantaneous Frequency of Oscillation
$-D_0$ = Uniform Delay		ω_c = Natural Frequency
$-\frac{d\Theta}{d\omega}$ = Delay Dispersion		$\frac{d\theta}{dt}$ = Frequency Modulation of the Impulse Response

definitions are in common usage. Others, though widely used, have vague meanings; these are given strict definitions. Still others are names given to quantities believed to have been previously unlabeled. These have been carefully chosen to be unique and descriptive.

The terms phase dispersion, phase slope, phase-slope dispersion, and delay dispersion will be unfamiliar to most readers. Phase dispersion is that part of the phase spectrum of a waveform (or the phase response of a network) which is varying nonlinearly with frequency. Phase slope is the frequency derivative of the phase spectrum of a waveform and is in a sense similar to the delay of a network.

Phase-slope dispersion is the frequency derivative of the phase dispersion of a waveform and is therefore that part of the phase slope that varies with frequency. Delay dispersion is the corresponding term for a network and is that part of the network delay (group delay is meant) that varies with frequency.

Because the shape of a waveform is completely specified by the amplitude spectrum and the phase dispertion, the expression for the spectrum can be simplified by omitting the constant and linear phase terms. Likewise, a network can be represented in simplified form by its amplitude response and phase dispersion. These simplifications are used often in the following explanations.

General Considerations of Pulse Compression

As stated previously, to achieve good range resolution and good velocity resolution simultaneously, a waveform with the required time duration and bandwidth must be found, and a means of generating the waveform and processing the received echoes must be devised. As approaches to solving these problems, a number of systems known as pulse-compression systems have been developed. In this section, pulse compression is considered in a very general way, and requirements of a waveform suitable for pulse compression are derived.

In a pulse-compression system, a long pulse having a time duration T seconds and a bandwidth F cycles, where $T \cdot F$ is greater than one, is somehow generated and transmitted. Received echoes are processed in such a way that short pulses having a time duration $1/F$ are obtained. The ratio of the duration of the long pulse to that of a short pulse is an important system parameter called the *compression ratio*. Writing the expression for the compression ratio, $K = T/(1/F) = T \cdot F$, it is seen to be the same as the previously defined figure of merit. Thus, the compression ratio can be used as a measure of the theoretical performance of a particular system.

To understand how pulse compression can occur, consider two pulses having identical amplitude spectra. The first has phase dispersion while the second has *no* phase dispersion. The time durations of these two pulses are to be compared.

The absence of phase dispersion in the spectrum of the second pulse means that at a particular instant (given by the uniform phase slope) all frequency components have the same phase. The magnitude of the envelope is, in general, given by the vector sum of all components. When the components are all in phase, the magnitude is the arithmetic sum of the components and is, at that instant, as large as it can ever be. The frequency components of the pulse with dispersion, however, are never all in phase, and the magnitude of its envelope, consequently, is never as great as that of the second pulse.

The energy of these pulses can be found in either of two ways: 1) the amplitude spectrum can be squared and integrated over all frequency or 2) the envelope can be squared and integrated over all time. Since the two pulses have identical amplitude spectra, they have the same energy. In order to satisfy this equality, the pulse with dispersion, which has the lesser magnitude, must persist over a longer time than the pulse without dispersion. The pulse without dispersion has the largest magnitude possible for the given amplitude spectrum. It is, therefore, the shortest pulse obtainable having that amplitude spectrum. The length of this pulse is, essentially, $1/F$.

It follows from the preceding argument that if the phase dispersion is removed from the spectrum of a pulse, the pulse will be shortened in duration, or compressed. Since the phase dispersion of the output waveform of a network is the sum of the phase dispersion of the network and that of the input waveform, to remove the phase dispersion of a pulse, therefore, a network having an identical but negative phase dispersion is needed.

Now, conditions for pulse compression to occur can be stated:

1) The long pulse to be compressed must have phase dispersion; *i.e.*, the phase spectrum must be a nonlinear function of frequency.
2) The long pulse must be passed through a network having phase dispersion which is equal and opposite to that of the pulse.

No specifications have so far been placed on the amplitude response of this filter. The amplitude response can, however, be determined by considerations other than those of pulse compression.

If a signal plus noise is passed through a filter, the signal-to-noise ratio at the output will be a maximum for a very particular filter. This filter has an amplitude response equal to the amplitude spectrum of the waveform and a phase dispersion which is equal and opposite to that of the waveform. The maximum signal-to-noise ratio is given simply by the ratio of the total waveform energy to the noise power per unit bandwidth. This de-

pendence of signal detectability on signal *energy* was stated at the beginning of this paper.

When the filter response and signal spectrum are related in the way just described, the filter is said to be matched (or North-matched)[2] to the waveform and is called a *matched filter*.

It is easy to see how the matched-filter conditions work to produce a high signal-to-noise ratio. It has already been shown that pulse-compression condition 2), *phase matching*, produces the largest pulse magnitude possible for a given amplitude spectrum. On the other hand, the noise components are randomly related in phase, and the phase dispersion of the filter has no effect on the noise amplitude. Also, setting the amplitude response of the filter equal to the amplitude spectrum of the pulse, *amplitude matching*, emphasizes the parts of the spectrum having higher signal-to-noise ratio and places less weight on the parts having less signal-to noise ratio.

In some instances, it may not be desirable to conform to the matched-filter conditions exactly. For example, the pulse shape obtained at the output of the matched filter may have, in addition to the main peak, several additional secondary peaks or "sidelobes." By departing somewhat from amplitude matching, however, the sidelobes can be reduced with respect to the main peak.

In the case of a waveform with rectangular amplitude response, the output of the matched filter has a sin x/x-shape in the time domain. This waveform is characterized by relatively large sidelobes ($\sim$13.5 db down). If a lower sidelobe level is desired, it is necessary to change the shape of the filter response from rectangular.

The reduction in the sidelobe level is of course not without a penalty. Lower sidelobe level has to be traded against lower signal-to-noise ratio and increased pulse width. To demonstrate this, compare, for instance, the (cosine)^2 shaped response with the ideal rectangular response: 1) Signal-to-noise ratio is reduced by 1.76 db. 2) Sidelobe level is -33 db as compared to -13.3 db. 3) Pulse width (distance of first zeros) is doubled.

The amplitude spectrum shaping (or weighting) giving least pulse widening for a given sidelobe level (or lowest sidelobe level for a given pulse width) is obtained by using Taylor weighting.[3] With a design goal of -33-db sidelobe level, for instance, a pulse widening by a factor of approximately 1.5 takes place.

As another example, it may be impractical to have exact matching in the case of Doppler-shifted signals, as this would require a separate matched filter for each Doppler shift. By accepting slight mismatches in amplitude and the resulting decrease in signal detectability, a filter matched for signals not Doppler shifted may give satisfactory results for Doppler-shifted signals as well.

The Selection of the Radar Signal Waveform

As shown above, an improvement in radar-system performance can be achieved by using a signal waveform satisfying certain requirements. These requirements are:

1) A product of bandwidth and duration greater than unity (for good resolution of both range and velocity),
2) A rectangular amplitude spectrum (for best range resolution),
3) A rectangular envelope (for best velocity resolution), and
4) Phase dispersion (that is, a nonlinear phase spectrum to permit pulse compression).

Although the requirements listed above eliminate a number of possibilities, a particular waveform has not yet been specified. By stating a fifth condition, one waveform is suggested. It is then shown that this waveform satisfies all five of the stated requirements.

To effect pulse compression, the received signals must be passed through a network having phase dispersion which is equal and opposite to that of the waveform. This condition has been called *phase matching*. When a network and waveform are phase matched, the waveform, in passing through the network, is stripped of its phase dispersion and left with a phase spectrum consisting of terms linear and constant with frequency. This action, as just shown, produces the pulse compression.

Generally, in a radar system the received signals will have suffered a Doppler frequency shift which is not known *a priori*, except that it is between certain limits. In the case of most waveforms, an array of parallel phase-matching networks, each designed for a particular Doppler shift, would have to be assembled in the receiver. Usually the received signals would not be aligned exactly with one network. Some means of interpolating intermediate Doppler shifts have to be provided. Obviously, a great saving in equipment complexity will result if a waveform can be found that will remain phase matched to *one* network regardless of Doppler shifts. This is the *fifth* requirement of the desired signal waveform.

Consider the simplest case of phase dispersion, phase varying as the square of frequency (parabolic phase dispersion). This can be expressed by $\Phi(\omega) = a(\omega - \omega_c)^2$. The phase dispersion of the phase-matching network must be $\theta(\omega) = -a(\omega - \omega_c)^2$. If a signal with phase dispersion $\Phi(\omega)$ is subjected to a Doppler shift[4] ω_D, its phase dis-

[2] D. O. North, "An Analysis of the Factors which Determine Signal/Noise Discrimination in Pulsed-Carrier Systems," RCA Labs., Princeton, N. J., Tech. Rept. PTR-6C; June, 1943.

[3] T. T. Taylor, "Design of line-source antennas for narrow beamwidth and low side lobes," IRE Trans. on Antennas and Propagation, vol. AP-3, pp. 16–28; January, 1955.

[4] The Doppler effect is assumed here to cause a uniform frequency shift of all components of the waveform spectrum. This is a reasonable assumption for a waveform having a time-bandwidth product smaller than 10^4 and targets with orbital velocities.

persion becomes $\Phi(\omega+\omega_D)=a(\omega-\omega_c+\omega_D)^2$. If this signal passes through the network phase matched to the zero-Doppler, or unshifted signal, the phase dispersion of the output is:

$$\Phi(\omega+\omega_D)+\theta(\omega)=a(\omega-\omega_c)^2+2a(\omega-\omega_c)\omega_D+a\omega_D{}^2-a(\omega-\omega_c)^2,$$

or

$$\Phi(\omega+\omega_D)+\theta(\omega)=2a(\omega-\omega_c)\omega_D+a\omega_D{}^2.$$

Obviously, the phase dispersion has been removed, *independently of the Doppler shift*. The reader can easily satisfy himself that the higher-order dispersion functions do not have this property.

Notice that the phase spectrum of the output waveform consists of a uniform phase $a\omega_D{}^2$ and a linear term $2a(\omega-\omega_c)\omega_D$. The linear term implies a uniform phase slope $2a\omega_D$ which is linearly related to the Doppler shift. It is later shown how this uniform phase slope can be used to provide a simple but effective means of measuring target velocity.

To obtain maximum signal detectability, the amplitude response of the phase-matching network must be identical to the amplitude spectrum of the waveform. Clearly, with a single network, this *amplitude matching* is impossible to achieve for all Doppler shifts. However, using a network amplitude matched for zero-Doppler signals (with rectangular amplitude response) the loss in detectability increases linearly with Doppler shifts. The loss is only 3 db at a shift of one-half the network bandwidth.

Another approach to allow for Doppler shifts would be to use a phase-matching network which has sufficient bandwidth to completely pass received signals having the maximum expected Doppler shifts. The received signals will, therefore, not be distorted or reduced by Doppler shifts, but since the excess bandwidth lets in more noise, signal-to-noise-ratio will be less than in the case of true matched-filter operation. Some of this loss can be recovered by the use of a properly matched post-detection filter.

A waveform with rectangular amplitude spectrum and parabolic phase dispersion thus satisfies requirements 2), 4), and 5). To determine whether the product of bandwidth and duration is greater than unity [requirement 1)] and the envelope is rectangular [requirement 3)], it is necessary to determine the corresponding time function describing the waveform. To do this, the inverse Fourier transform is performed. The result is a waveform with a very nearly rectangular envelope and very nearly parabolic phase modulation. Furthermore, the time duration is given by $T=4\pi aF$, where F is the bandwidth and a is the coefficient of the phase dispersion $a(\omega-\omega_c)^2$. Given a particular bandwidth, the duration can be made as long as desired (subject to practical limitations) by using the proper phase dispersion. For convenience in further discussion, this waveform will be referred to as the LInearly Frequency MOdulated Pulse or LIFMOP waveform. This name follows from the fact that the frequency modulation, the derivative of the phase modulation, is very nearly linear.

The phase dispersion of the LIFMOP waveform has been expressed by $\Phi=a(\omega-\omega_c)^2$. The phase-slope dispersion, therefore, is $\sigma=-2a(\omega-\omega_c)$. Over the bandwidth of the waveform $2\pi F$, the phase-slope dispersion changes by $4\pi aF$, which is numerically equal to T, the duration of the waveform.

The significant term in the phase modulation of the LIFMOP waveform is $\phi=-t^2/4a$. The corresponding frequency modulation is $d\phi/dt=-t/2a$. Thus, the instantaneous frequency of the waveform changes by an amount $T/2a=4\pi aF/2a=2\pi F$, the bandwidth, between the beginning and end of the pulse.

Thus, the signal waveform with rectangular amplitude spectrum and parabolic phase dispersion (the LIFMOP waveform) satisfies all of the requirements for good radar performance:

1) Bandwidth-duration product greater than unity,
2) Rectangular amplitude spectrum,
3) Rectangular envelope,
4) Phase dispersion,
5) Single-filter phase matching.

By using this waveform in a radar system, high resolution of both range and velocity can be obtained. Pulse compression will occur for a wide range of Doppler frequency shifts using only one processing filter. In addition, the response of the filter to Doppler-shifted signals can provide a simple means of measuring target velocity.

Fig. 1 summarizes the time and frequency characteristics of the LIFMOP waveform. The time-domain characteristics indicate only the general form. The irregularities in the phase- and frequency-modulation plots are exaggerated to make them noticeable.

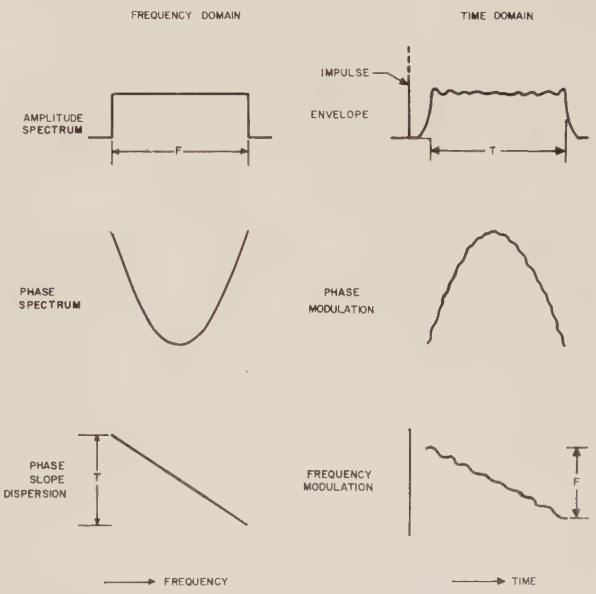


Fig. 1—Characteristics of the LIFMOP waveform.

A System to Use the LIFMOP Waveform

To obtain improved radar performance, not only must a waveform with suitable properties be found, but a system to use the waveform effectively must be devised. So far, the properties of the LIFMOP waveform have been shown to be consistent with good performance. Now, a system which utilizes the LIFMOP waveform to achieve this performance is outlined.

The first problem in devising a system is generating the waveform. One method of generation would be simultaneous amplitude- and phase-modulation of an oscillator. This method is called *active generation*. In view of the complexity of both the amplitude and phase modulation, however, active generation of the LIFMOP waveform is not practicable.

A more practical way of generating the waveform is based on the fact that the impulse response of a network is the inverse Fourier transform of the frequency response of the network. Thus, to generate a certain waveform, a network having a response identical to the waveform spectrum is driven by an impulse. To generate the LIFMOP waveform, a network is required that has a rectangular amplitude response and parabolic phase dispersion identical to the amplitude spectrum and phase dispersion of the waveform. If this network (called a LIFMOP filter) is excited by an impulse, the output will be the LIFMOP waveform. This method of generation is called *passive generation*.

Fortunately, an impulse in the mathematical sense, which is physically unrealizable, is not required for passive generation. All that is needed is a short pulse having a spectrum wide enough that over a band corresponding to the desired waveform spectrum, the amplitude spectrum is uniform and the phase spectrum is linear with frequency. A pulse like this is easily obtained, for example, from a blocking oscillator. In the following discussion, impulse will mean such a short pulse.

After the LIFMOP waveform is generated, it is heterodyned to a suitable center frequency, amplified, and transmitted. Received echoes are heterodyned to the original frequency band and amplified. To effect pulse compression, the received signals must be passed through a phase-matching network. Since this network has phase dispersion equal and opposite to that of the waveform, the phase dispersion of the received signals is removed and pulse compression takes place.

In Fig. 2, a simplified diagram of a pulse-compression radar system, filter 1 is the waveform generating network, and filter 2 is the phase-matching or pulse-compressing network. The characteristics of both filters are shown in Fig. 3. Here, instead of phase dispersion, delay dispersion is shown. Delay dispersion is the frequency derivative of the phase dispersion of a network. Since the phase dispersion of these networks is parabolic with frequency, the delay dispersion is linear with frequency. The amplitude response and delay dispersion of filter 1 are the amplitude spectrum and phase-slope dispersion of the transmitted LIFMOP waveform. In accordance with the pulse-compression requirement, the delay dispersion of filter 2 has a slope equal and opposite to that of filter 1. The impulse response of filter 1, which is, of course, the LIFMOP waveform, is shown in Fig. 1.

Before going further, the reader may wish to refer to Tables I and II to review the definitions of phase slope, delay dispersion, phase-slope dispersion, etc.

The operation of the simplified system of Fig. 2 can be explained with the aid of Fig. 4. Here, the delays of filters 1 and 2, and the phase slope of waveforms at various points in the system are plotted as functions of frequency. Fig. 4 (a) shows the case of a stationary target, and (b) and (c) apply to a moving target.

The explanation starts with the impulse, which has a phase slope uniform with frequency. A uniform phase slope means a time shift of a waveform from the time origin. In this explanation, the occurrence of the impulse is taken as the time origin; hence, the phase slope of the impulse is zero.

The phase slope of the output of a network is the sum

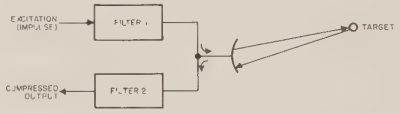


Fig. 2—Pulse-compression system using separate generating filter 1 and processing filter 2.

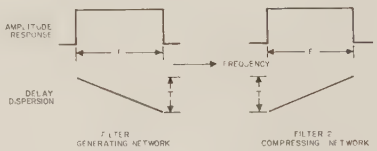


Fig. 3—Characteristics of generating and processing filters.

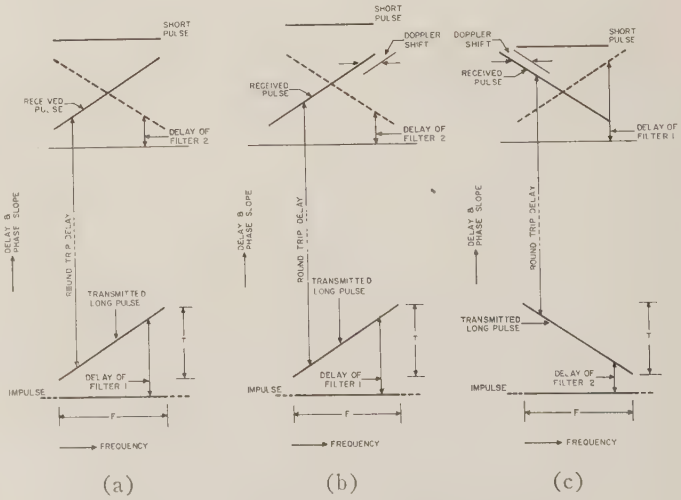


Fig. 4—Graphical explanation of the operation of a LIFMOP system.

of the input phase slope and the network delay. The phase slope of the long pulse that is transmitted (the impulse response of filter 1) is, therefore, the delay of filter 1. Of course, filter 1 passes only that part of the impulse spectrum included in the filter pass band (width F).

The time required for the signal to travel to the target and back to the radar is equivalent, in the frequency domain, to a uniform delay. If no Doppler shift occurs, the case shown in Fig. 4(a), the phase slope of the received long pulse is the sum of a uniform delay and the phase slope of the transmitted pulse. The received pulse then passes through filter 2, and the filter delay is added to its phase slope. Since filter 2 is phase matched to the long pulse, the phase dispersion is removed, *i.e.*, the phase slope of the output pulse is uniform. Thus, the output spectrum is flat in amplitude over the band F and has no phase dispersion. Such a waveform (a band-limited impulse) has a sin x/x shape, and a duration between first zeros of $2/F$. This short pulse occurs at a time given by the magnitude of its uniform phase slope (which is the sum of all the network and propagation delays).

Fig. 4(b) illustrates what happens when the target is moving (in the case shown, toward the radar). Here, not only is the received pulse delayed by the propagation time to the target and back, but because of the Doppler effect, the frequency of each component of the received signal is shifted (in the case shown, increased) an amount proportional to the target radial velocity. Now, when the delay of filter 2 is added to the phase slope of the received signals, the phase slope of the output signal is again uniform but greater than in the zero-Doppler case. The short pulse appears, therefore, at a later time than for a nonmoving target at the same range. Because the relation between phase slope of the long pulse and frequency is linear, this time shift is proportional to the Doppler shift, and thus to the target's radial velocity.

Since the short-pulse position depends on both target range and radial velocity, the simple system of Fig. 2 gives ambiguous results. The ambiguity can be resolved, however, by interchanging the positions of filter 1 and filter 2. Fig. 4(c) shows what happens when filter 2 is used as the generating filter and filter 1 is used as the compressing filter. For the same target range and velocity assumed in Fig. 4(b), the short pulse is again shifted in time by an amount proportional to radial velocity, but now in the *opposite* direction, that is, earlier.

By alternating the positions of the two filters, say on every other pulse, a target will produce short pulses at two time positions. Displayed on an A-scope, the return from a typical target will appear as shown in Fig. 5. The center of the two pulse positions is a measure of the target range, while the separation is proportional to the target radial velocity. If the Doppler frequency shift is f_D, the time separation will be $2Tf_D/F$.

A filter with rectangular amplitude response and parabolic phase dispersion (the LIFMOP filter) is a complex device. A significant reduction in system complexity is obtained by using one filter for both generation and compression. Such a single-filter system is made possible by the technique of *sideband inversion.*

When a waveform is sideband inverted, its amplitude spectrum is, in effect, inverted about the center frequency. The phase spectrum is likewise inverted but, in addition, a change of sign occurs. Thus, if the phase slope increased with frequency before sideband inversion, it will decrease with frequency after sideband inversion. When the LIFMOP waveform is sideband inverted, because of symmetry, the only change that occurs is a change of sign of the phase-dispersion coefficient.

Fig. 6 shows a simplified block diagram of a system using a single filter in conjunction with a *sideband inverter.* The path of the signal through the system for a complete cycle of operation (two pulse periods) is as follows:

The impulse enters the LIFMOP filter. The resulting output is a signal with phase slope decreasing with frequency. Since switch S3 is in position *a*, this signal proceeds to the transmitter and is radiated.

The received signal reaches switch S1 and is routed to the input of the sideband inverter. It leaves the sideband inverter with phase slope increasing with frequency. By way of S2, this inverted signal is fed to the input of the LIFMOP filter. Conditions are identical to those shown in Fig. 4(b) and at the output of the LIFMOP filter a short pulse is obtained.

At the beginning of the next period, another long

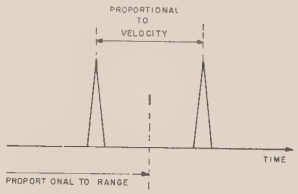


Fig. 5—Typical processed return from a moving target (A-scope display).

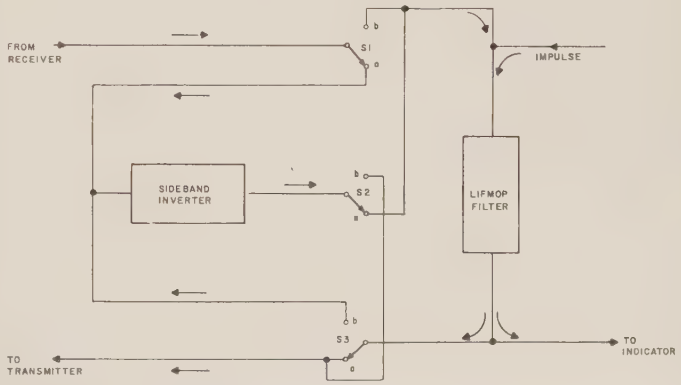


Fig. 6—Simplified block diagram of a LIFMOP system.

pulse is generated by the LIFMOP filter. The switches S1, S2 and S3 are now thrown to position *b*. The long pulse from the LIFMOP filter is, therefore, routed to the sideband inverter, and thus a signal with phase slope increasing with frequency is transmitted.

Upon reception, this signal is passed through S1-*b* directly to the input of the LIFMOP filter. Conditions are now identical to those in Fig. 4(c) and again a short pulse is produced.

For reasons of clarity in explanation, the sideband inverter is shown here as a separate block of the system. The sideband inverting operation can, however, be conveniently performed in conjunction with the heterodyning which usually takes place in any radar system. It is only necessary to use two different local oscillators in the transmit and receive paths of the system. Interchanging these local oscillators by suitable gating circuits on consecutive pulse periods provides sideband inversion alternately before transmission and after reception.

Conclusion

The limitations on radar performance imposed by the simple radar pulse can be overcome by use of the signal waveform with rectangular amplitude spectrum and parabolic phase spectrum (the LIFMOP waveform). This waveform very closely meets the requirements for good radar performance, namely large duration-bandwidth product, rectangular amplitude spectrum, and rectangular envelope. Moreover, a parabolic phase spectrum has the unique feature that waveform and compressing filter remain phase matched regardless of Doppler frequency shifts. Furthermore, the LIFMOP waveform is a practical signal waveform since a radar system using it can be easily devised and duration-bandwidth products over 100 readily implemented.

Acknowledgment

The authors wish to acknowledge the assistance of R. V. Lang and O. B. Waddell of the General Electric Legal Operation in making release of this paper possible.

Airborne Pulse-Doppler Radar*

L. P. GOETZ†, MEMBER, IRE, AND J. D. ALBRIGHT†, MEMBER, IRE

Summary—**Doppler radars are employed for the detection of moving targets whose radar echo area is much smaller than the ground clutter return. Moving targets are separated from clutter on a frequency basis by utilizing the Doppler phenomenon. Continuous-wave Doppler radars have a practical maximum-range capability because the leak-through between the transmitter and receiver causes receiver saturation. This limitation is overcome in pulse-Doppler radar by time-sharing the transmitting and receiving cycles. This paper discusses a typical pulse-Doppler radar and the basic design considerations for the selection of the pulse-recurrence frefrequency, elimination of clutter, range determination, and oscillator stability requirements.**

Introduction

IN the past few years, standard-pulse radar has reached a high state of development. Even with its relative perfection, pulse radars in airborne applications have limited detection capability for targets in the presence of ground clutter. One method of separating moving targets from ground clutter is to take advantage of the Doppler shift in the carrier frequency associated with a moving target. Continuous-wave (CW) Doppler radars[1,2] have an ultimate range capability, which is limited by receiver saturation, caused by transmitter leak-through. For airborne applications, where a single antenna is essential, this limitation is overcome by time-sharing the transmitting and receiving cycles. Radars of this class are described as pulse-Doppler (PD) radars. A PD radar differs fundamentally from standard-pulse radars, and parameter selection and mechanization must be made according to a new set of concepts.

It is the purpose of this paper to describe a typical airborne PD radar and to present a condensed review of PD radar design considerations in those areas which experience has shown to be the most important from a systems design point of view. These areas are the selection of the pulse-recurrence frequency (PRF), ground clutter considerations, and oscillator stability requirements. In addition, a section has been included on the determination of range performance.

* Received by the PGMIL, January 17, 1961.
† Air Arm Div., Westinghouse Electric Corp., Baltimore, Md.

[1] L. N. Ridenour, "Radar System Engineering," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1; 1947.

[2] R. S. Sargent, "Moving target detection by pulse doppler radar," *Electronics*, vol. 27, pp. 138–141; September, 1954.

A Pulse-Doppler Radar

A brief discussion of a typical PD radar will be given in this section. The considerations which lead to this mechanization will be discussed later. The block diagram of this PD radar is shown in Fig. 1. The block diagram is of a PD radar which must search (S) for its target, and then track (T) the target in velocity, range, and angle. In search, the target is only located in velocity and angle.

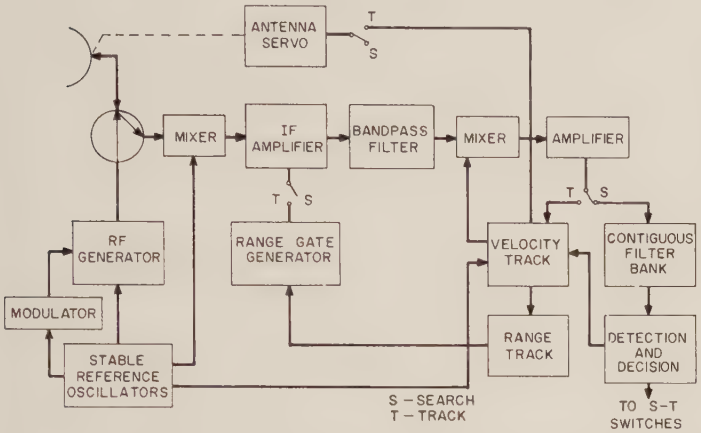


Fig. 1—Block diagram of typical pulse-Doppler radar.

The Doppler shift produced by a practical moving target is only a small per cent of the transmitted frequency.[1] In order to detect this small shift, highly stable transmitters and local oscillators are required. Typically, a stable reference source is used for radio-frequency (RF) generation, and a fixed off-set from this frequency is used for the local oscilator. This technique simplifies stable oscillator design, and the tolerable short-term stability. It is also beneficial if all oscillators used in the radar have a common reference, as indicated in the block diagram.

The modulator keys the transmitter at the high-pulse recurrence frequency (PRF) required in a PD radar. Typically this is the in order of one hundred kilocycles. A return signal is intercepted by the same antenna as used for transmission, and is routed to a mixer via an isolator. The mixer translates the received frequency to an appropriate intermediate frequency (IF). The IF signal is amplified and passed through a band-pass filter, which selects an appropriate sideband for target tracking. Filters for rejection of main-beam and altitude-line clutter may also be included at this point in order to obtain additional clutter attenuation. The reduction of bandwidth in the receiver as early as possible after range gating simplifies amplifier design by removing the clutter and reducing the noise power, hence reducing amplifier dynamic range requirements.

The output of the band-limited IF feeds another mixer which translates the spectrum to one satisfactory for construction of a contiguous filter bank. This filter bank encompasses the target Doppler frequency range of interest. In order to prevent noise folding and ambiguous velocity problems, the contiguous bank must be so designed that it does not encompass more than one octave of frequency. The filters of the contiguous filter bank normally terminate in integrators. These integrators improve the system range performance and simplify readout.[3,4]

When a target is detected, the detection and decision unit switches the system to the track mode. In track, the antenna is stopped, the velocity track loop is locked on target, using information from the contiguous filter bank for oscillator positioning, range track is initiated, and angle track is started.

Techniques similar to those used in pulse radar are satisfactory for range and angle track, while velocity track is performed with either a phase or frequency track loop. It is common PD radar practice to feed the velocity information to the range-track feedback loops to aid in target range tracking.

Clutter and Target Spectra

Clutter in a PD radar set occurs when a non-target object is seen either in the main beam or antenna sidelobes. The clutter will appear to have the general character of noise in a band adjacent to a spectral line and, depending on the size of the object, its bearing, relative velocity and range, it may obscure a desired target. The basic problem is to arrange the parameters of the radar set such that clutter is kept to a minimum in the region where targets may be expected.

When a PD radar set is flown in an airplane, clutter can come from three sources, as illustrated in Fig. 2. Main-beam clutter comes from weather effects and from the ground if the main beam dips below the horizon. Sidelobe clutter comes from the antenna sidelobes striking the ground at all angles giving Doppler shifts both above and below the transmitted frequency. This produces a range of Doppler frequencies with a spread corresponding to twice the ground velocity of the aircraft. At any instant, one of the sidelobes will be looking straight down at the earth and will return a signal with zero Doppler shift. This is identical to the altitude line in ordinary pulse radar.

Fig. 3 shows how the clutter spectrum is distributed about each line in the return spectrum. To appear in a clutter-free, region a target must have a closing velocity

[3] N. Weinberg, J. M. Walker, and R. J. Metz, "Rapid detection of coherent signals in noise," 1960 IRE International Convention Record, pt. 8, pp. 222–232.

[4] S. A. Meltzer and S. Thaler, "Detection range predictions for pulse Doppler radar," 1960 IRE International Convention Record, pt. 4, pp. 105–115. This development does not consider eclipsing losses.

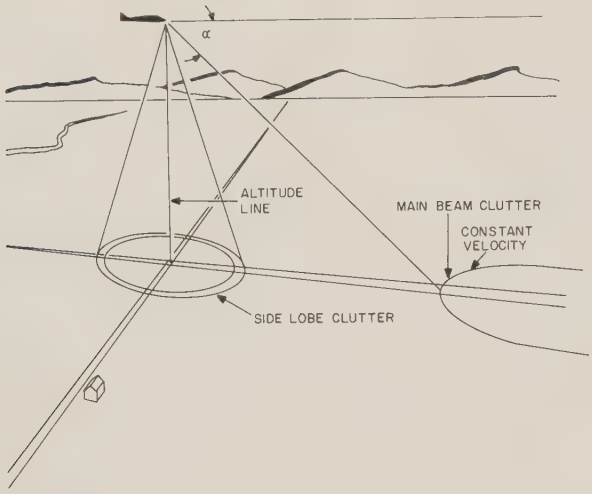


Fig. 2—Sources of clutter in an airborne pulse-Doppler radar.

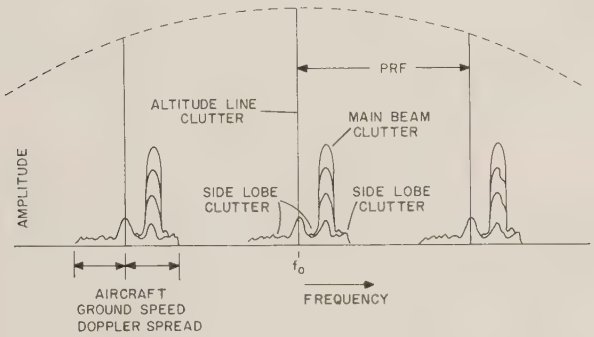


Fig. 3—Typical appearance of clutter around each spectral line in an airborne nose-mounted pulse-Doppler radar.

such that its Doppler shift is greater than the closing clutter to the right of the spectral lines in Fig. 3. Also, the closing velocity cannot be too great or it will run the Doppler shift into the opening clutter of the next spectral line. Thus, there is a limit on the range of closing target velocities which can be handled unambiguously by the radar. In a similar manner, opening targets are also limited.

The effect of each of the various types of clutter for different target headings is shown in Fig. 4. An aircraft with a scanning antenna is flying in the direction shown. Two targets are indicated, one directly ahead of the aircraft and one at an azimuth angle of 45°. Each target is located at the center of the circles. The three circles, labeled 1, 2, and 3, represent particular ratios of target-ground velocity to interceptor-ground velocity. A target vector showing velocity ratio and direction may be drawn from the center of each target position. The tip of the vector will show whether the target is in the clear or is obscured by clutter. In Fig. 4 the two target vectors show the targets to be outside the clutter. If both targets were traveling in the opposite direction, the one straight ahead would be obscured by opening sidelobe

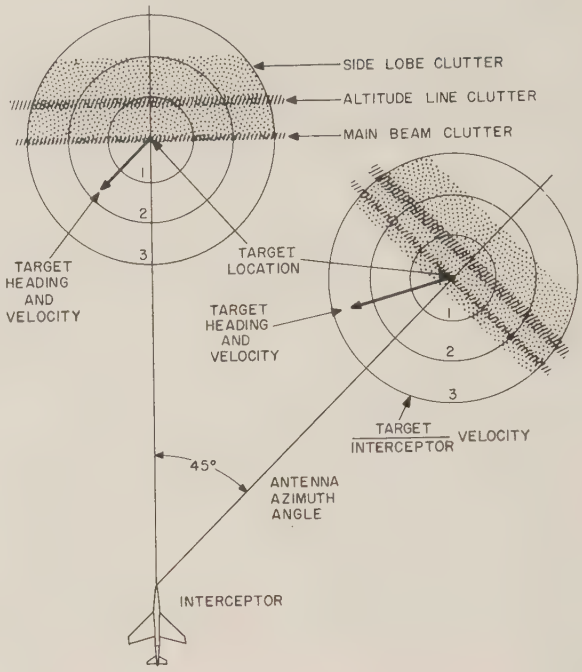


Fig. 4—Clutter and target vector diagram.

clutter, while the target on the 45° bearing, because of its higher velocity, would be clear of opening clutter.

Selected Spectrum

The returned signal will be essentially the same as the transmitted signal, with the exception that each spectral line will be shifted by the amount of Doppler frequency f_d. The magnitude of this Doppler shift is a measure of the relative velocity between the radar set and target. If this shift could be measured directly, it would provide within itself all the information on the target measurable on a frequency basis. As it is not practical to measure the Doppler shift directly, the received signal is heterodyned to a lower frequency. After filtering, the Doppler information is contained in the frequency of the remaining sidebands.

The spectral lines in the transmitted pulse are each shifted by the Doppler phenomenon when reflected from moving targets. Each interspectral line Doppler pattern contains all the available information, and therefore any one could be used. For any reasonable transmitter pulse shape, the maximum spectral amplitudes are adjacent to the center line. To obtain the greatest received amplitude and to simplify the local oscillator requirements, either Doppler band adjacent to the center spectral line is used.

If the total energy in the transmitted pulse were reflected by a single target moving with respect to the radar, the entire returned spectrum would be shifted by the Doppler frequency f_d. If the transmitted carrier is f_0, the input to the mixer is then f_0+f_d (for range clos-

ing) and f_{LO}. Spectra are generated at $f_{IF}+f_d$, $2f_0+f_{IF}+f_d$ and higher harmonics of f_0. Of these, only $f_{IF}+f_d$ is generally available. A segment of the output spectrum is shown in Fig. 5.

After suitable amplification, the spectrum may be video detected. The maximum amplitude of the video spectrum is at zero frequency, and that half of the *IF* spectra below the center frequency is folded over onto the upper half. Frequencies such as $f_{IF}-f_r+f_d$ in Fig. 6 become equal to f_r-f_d in the video spectrum.

If the Doppler shift exceeded one half the PRF, the spectral lines will cross, and it will be very difficult to resolve the Doppler frequency. Therefore, for a maximum relative velocity, the following relation must be satisfied:

$$f_{d\,\max} \leq \frac{f_r}{2}. \tag{1}$$

Noise is folded along with signal spectral lines. As the noise power is noncoherent from one side of the IF spectra to the other, that in the lower half of the spectrum adds directly to that in the upper half. The noise power in the video spectrum is thus doubled, and the S/N is reduced by a factor of 2 compared to the IF S/N. For this reason, direct video detection at this point is undesirable.

Single-Sideband Operation

To avoid both the confusion resulting when $f_d=f_r/2$ and the reduced S/N due to folding, the most obvious corrective action is to remove by IF filtering all frequencies on one side of f_{IF}. As repeated Doppler shift patterns associated with higher-order spectral lines are redundant, they, along with their associated noise, may be filtered out. Thus, if the IF bandwidth is narrowed to approximately f_r, no Doppler information is lost and an improved S/N is produced.

Minimum PRF

Fig. 6 illustrates a section out of the IF spectrum with Doppler lines near f_{IF}, $f_{IF}+f_r$ and $f_{IF}-f_r$. In the case of an airborne radar set, the return Doppler signals are referenced to ground velocity, f_{d_A} representing the maximum ground velocity Doppler of the aircraft and f_{d_T} the maximum ground velocity Doppler of the target. In the case illustrated, the radar-carrying aircraft is assumed to have a ground velocity one-third that of the target. The selected spectrum is the part of the total Doppler return extracted by a band-pass filter for determining the Doppler shift. Both closing ($+f_{d_T}$) and opening ($-f_{d_T}$) aspects are assumed for the target.

In order to prevent Doppler signals from adjacent PRF spectrum lines overlapping in frequency and thus being confused, f_r must have a minimum spacing from

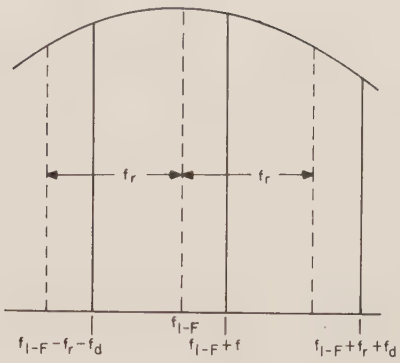


Fig. 5—A section of the IF spectrum output from the first mixer.

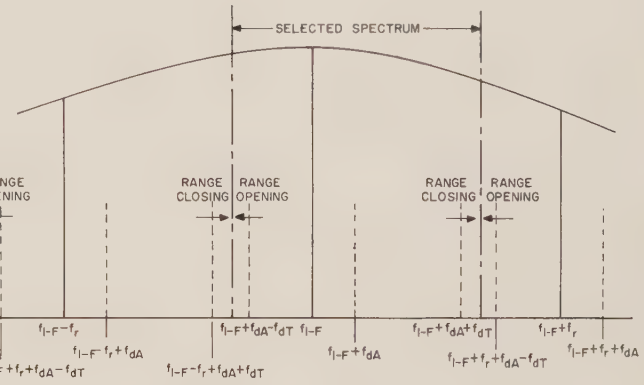


Fig. 6—Spectrum passed by SSB filter located so that both opening and closing targets ($\pm f_{dT}$) are obtained. Conditions shown for nose-mounted forward scanning radars.

f_{IF}. Fig. 6 shows that

$$f_{r\,\min} \geq f_{IF} + f_{d_A} + f_{d_T} - (f_{IF} + f_{d_A} - f_{d_T}) = 2f_{d_T}. \tag{2}$$

This relation infers that the Doppler-detection filter bank is constructed such that the center of the filter bank is at the ground velocity of the radar-carrying aircraft. This detection system requires a knowledge of the aircraft-ground velocity and a tracking device to keep the filter bank centered at the interceptor ground velocity.

If the Doppler filter bank is fixed in frequency, then the filter bank must be wider than the previous arrangement in order to account for all aircraft-ground velocities. In order to prevent false Doppler indications for this case, (2) is modified to

$$f_{r\,\min}' \geq 2f_{d_T} + f_{d_A}. \tag{3}$$

Eqs. (2) and (3) illustrate the advantage, as far as PRF selection is concerned, of a mechanization in which the target detection filters are placed in such a manner that they detect a fixed maximum velocity ($\pm f_{d_T}$) above and below the interceptor-ground velocity (f_{d_A}). In addition, fewer filters are required in the first system.

A further reduction in the PRF can be accomplished for a low-speed interceptor if only head-on attacks are

required. In this case, it is only necessary to border the Doppler region with a safety region so that the filter banks do not detect in the sidelobe clutter region associated with $f_{IF}+f_r$. For this case

$$f_{r\,\min}'' \geq f_{d_T} + 2f_{d_A}. \tag{4}$$

Unambiguous velocity can be determined from ambiguous velocity measurements if the PRF is switched. In this case, all closing targets will appear in the same filter as the PRF is switched, whereas all opening targets will appear in a different filter with a change in PRF. Thus, by observing a target in at least two PRF's, it can be determined whether the range is opening or closing.

This combines the advantages of the previous cases. The lower minimum PRF and hence, lower number of filters for a multiple PRF condition, are incorporated with the ability to detect head-on closing and rear hemisphere-aspect targets as noted in (4).

Unambiguous Range Resolution

In a PD radar, as in a conventional pulsed radar, the range to a target is proportional to the time interval from the transmitted pulse to the reception of the return echo. The high PRF employed by PD radar, however, causes many-time-around echoes and corresponding ambiguities in range information. To resolve such ambiguities, several techniques are possible.

The first techniques to be suggested are identical to those used in CW Doppler radar. These involve the modulation of the carrier. For example, if the transmitted wave is frequency or amplitude modulated, the return Doppler-shifted signal will be modulated in the same manner. As there is a time delay between the received and transmitted signals, the measurement of the phase shift or frequency translation of the return signal will determine the range to the target. In theory, such methods might be applied, but there are certain practical disadvantages in their direct application to PD radar.

First consider the application of pulse modulation for ranging to a PD radar. There are four basic types of pulse modulation,[5] namely, pulse-position, pulse-duration, pulse-frequency, and pulse-amplitude modulation. If a sinewave is employed as the modulation,[6] it can be shown that multiples of this modulation frequency will appear above and below all the carrier spectra lines. This means that the clutter will also be smeared in the same manner. The net result is that the clutter can occupy the entire pulse spectrum.

Another ranging system is to linearly frequency-modulate the carrier and the local oscillator. Assume that the target has been located and has a known Doppler shift. The Doppler shift will be changed when the linear frequency modulation is applied, the amount of the shift depending upon the rate of the frequency modulation, the true range of the target, and whether the frequency-shift is increased or decreased from the nominal carrier. Knowledge of the two Doppler frequencies and the modulation rate permits the determination of true range. Clutter will also be smeared in this scheme.

A ranging system, which overcomes the limitations outlined above, uses multiple PRF's.[7] The principle of the multiple PRF ranging system is most easily understood by considering a two-PRF system. Fig. 7 illustrates such a system. The basic ranging frequency is shown, and two higher (PRF) frequencies are derived from the basic frequency by multiplying by two integers. In the case illustrated, the integers are 3 and 4 for simplicity of illustration. The true target range corresponds to a target return in the second transmitter interpulse interval. The ambiguous target returns are shown at each of the two PRF's. It is seen that the ambiguous range gates coincide at the true range. This technique can be extended to an infinite number of PRF's, and true range determined by multiple coincidences. In practice, however, a two- or three-PRF system can be used to resolve most range problems.

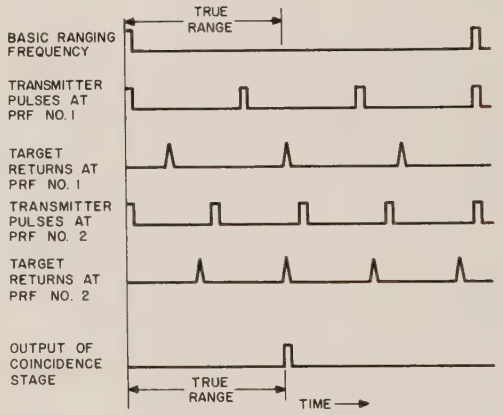


Fig. 7—A two-PRF ranging system.

There is a disadvantage to a multiple-PRF ranging system when multiple targets at the same velocity are detected in an antenna beamwidth. In the simple case of two targets and a two-PRF ranging system, there are four possible solutions. The relation is

$$S = M^Q \tag{5}$$

where S is the number of solutions, M is the number of targets, and Q is the number of PRF's.

The previous section can be summarized by stating that, where true range is required in a PD radar application, a careful investigation must be made to find the optimum ranging system for that application. As noted

[5] "Reference Data for Radio Engineers," International Telephone and Telegraph Corp., New York, N. Y., 4th ed., ch. 19, pp. 546–547; 1956.

[6] It is not necessary that the modulation be a sinewave; for example, noise could be employed.

[7] W. A. Skillman and D. H. Mooney, "Multiple High-PRF Ranging," presented at the Natl. Convention on Military Electronics, Washington, D. C.; June 27–29, 1960.

by Seibert,[8] velocity must always be determined in a PD radar before true range can be found.

Range Calculations

Range calculations for a PD radar are similar to the range calculation of other types of radar with one exception, that of eclipsing by the transmitter. A method for taking eclipsing into account is described in this section.

It is shown in Appendix I that the idealized range of a PD radar that uses the Doppler shift of the principal PRF spectral line is

$$R_0^4 = \frac{P_t d^2 G^2 \lambda^2 \sigma}{(4\pi)^3 k T \Delta F \overline{N F} d_n \delta}. \tag{6}$$

Holding all parameters in (6) fixed, except d, the range equation becomes

$$R_0 = C\sqrt{d}. \tag{7}$$

This is the case if a single fixed-range gate is used for target detection and the target is not eclipsed. It can easily be shown that a single fixed gate covering the entire interpulse period is equivalent to a narrower gate that is swept or searched over the same period, assuming the same antenna speed in both cases. The former is to be preferred for ease of mechanization and prevention of clutter modulation due to range-gate motion.

The method of Marcum[9] is equivalent to a CW Doppler system as far as range performance is concerned. The equation for Marcum's n is the same as for the latter system:

$$n = T_{fa}\Delta F. \tag{8}$$

In search, when the detected signal from the velocity channels is integrated, the number of signal varients that are integrated is N. N is determined by the equation

$$N = \frac{\theta \cdot \Delta f}{\omega_{\max}}. \tag{9}$$

The true detection curve may now be constructed by combining Marcum's or Swerling's curves for the proper n, N and R_0 with (6). This equation is modified to take target eclipsing into account by recognizing d as an effective duty cycle. This is shown in Fig. 8 for a simple case where the PRF is much lower than in practice. The first diagram shows the transmitted energy as a function of time. Fig. 8 shows the ratio of the duty cycle of the signal to the transmitted duty cycle as a function of the range of the target. For instance, if a target is at a range corresponding to T, it is completely eclipsed and the ratio of the duty cycle is zero. When the target moves to $T+\tau$, it is completely clear and the ratio becomes unity. To find the true R_0, R_{0_e}, corresponding to a particular position of the target, the square root of the duty cycle is taken. This has been done in Fig. 8(c). Now a series of detection curves for various values of R_{0_e} can be drawn, as in Fig. 8(d). The probability of detection at each range is found by reference to the particular curve that is applicable to the value of R_{0_e} at that range. The intersection of the dotted lines in Fig. 8(d) with the appropriate detection curves are points on the true detection curve which may then be sketched in, and is reproduced alone for greater clarity in Fig. 8(e). This is then the true detection curve when eclipsing is taken into account. In the actual radar, the PRF is much higher than shown and many more notches appear in the curve. This makes the drawing of the curve quite laborious, since each notch has a different shape and more of the original curves must be drawn to determine this shape.

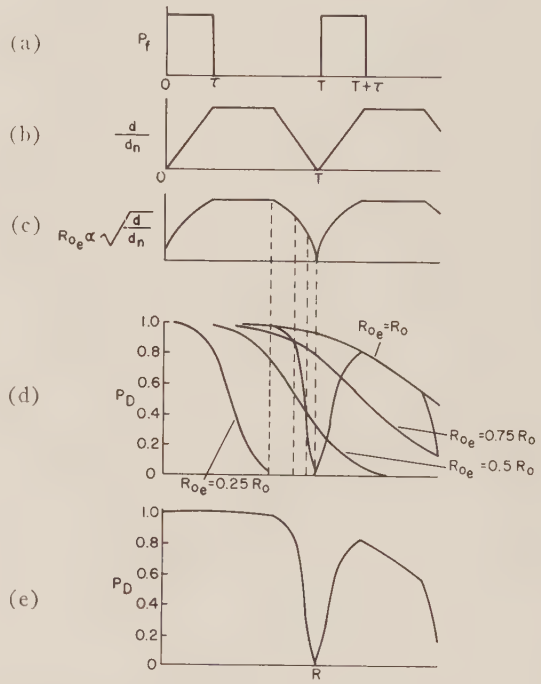


Fig. 8—Generation of detection probability curves.

The next step is to construct a curve showing the cumulative probability of detection for a target approaching the radar at a constant velocity. This curve has:

$$P_c = 1 - \sum_{d_n=1}^{N} (1 - P_{d_n}). \tag{10}$$

The target is usually assumed to start at a range where the probability of detection is small, and to move radially at a uniform speed. Since it is not known beforehand at what range the target appears in the scan pattern, a set of cumulative curves are determined for dif-

[8] W. Seibert, "A radar detection philosophy," IRE Trans. on Information Theory, vol. IT-2, pp. 204–221; September, 1956.

[9] J. I. Marcum and P. Swerling, "Studies of target detection by pulsed radar," IRE Trans. on Information Theory, vol. IT-6, pp. 59–144; April, 1960.

ferent initial positions within the range interval ΔR corresponding to the distance closed in a frame time. The values of P_c within each ΔR interval are averaged to obtain the total cumulative probability P_{t_c}. A new curve may then be drawn using the value of P_{t_c} at the middle of each ΔR interval.

In the early stages of system design, it is desirable to be able to compute the range performance with as little labor as possible. Normally, the cumulative detection curve is as good as the total cumulative curve for this purpose. However, with the true detection curve, the total cumulative curve must be computed because the cumulative curves bear little resemblance to the total cumulative curve. To alleviate this problem, it is found that approximately the same total cumulative probability curve is arrived at if the averaging process takes place before cumulation rather than after. This is easier to do, since the shape of the average curve, called the smoothed detection curve, is essentially independent of the PRF and is, of course, much easier to plot than the true curve. The construction of the smoothed curve is shown in Fig. 9. The probability of detection is averaged over each "loop" of the true curve to obtain one point on the smoothed curve, which may then be drawn through the points for all the loops. Only a limited number of points are needed to determine a smooth curve, which may then be used to construct a single cumulative curve. This curve is the total cumulative probability curve, since the averaging over initial positions has already been done. The curve is slightly affected by the initial position chosen, but the shift in range at high probability of detection is quite small.

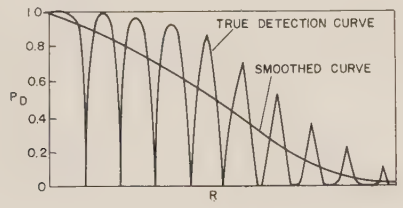


Fig. 9—Smoothed detection curve.

Search Mechanization Considerations

In Appendix I it is shown that for an idealized PD radar that searches a fixed volume of space in a fixed time, the idealized range R_0 can be expressed as

$$R_0 \propto (P_a^{1/4} D^{1/2}). \tag{11}$$

In order to derive this equation, the following assumptions were made:

1) The system sensitivity (noise figure) was fixed.
2) The velocity resolution was fixed.
3) A contiguous filter bank was employed to cover the Doppler range of interest.
4) The system could receive energy 100 per cent of the time; that is, eclipsing was not considered.

In this section, some of the practical constraints which must be considered in a PD search mechanization will be introduced.

Consider first an elementary PD radar which must search in range and velocity simultaneously. If the transmitter has a duty cycle d, then $1/d$ range intervals must be searched. Let the total Doppler frequency to be searched be F_v, and search this interval with a filter having a bandwidth Δf. The total number of velocity intervals is then $F_v/\Delta f$. The total number of intervals to be searched, P, in both velocity and range is

$$P = \frac{1}{d} \cdot \frac{F_v}{\Delta f}. \tag{12}$$

What is really of interest is the total time T_t required to search for a target. The build-up time of a band-pass filter is approximately $1/\Delta f$. Incorporating this relation in (12), the total time becomes

$$T_t = \frac{1}{d} \cdot \frac{F_v}{(\Delta f)^2}. \tag{13}$$

When a contiguous filter bank is used for detection, the search time T_t is reduced by the number of filters. This is the reason contiguous filters are generally used in PD radars in search operations. The idealized search derivation considered the use of a contiguous filter bank, but not the influence of the duty cycle.

The influence of the duty cycle on the search time in a practical system is actually much more complex than indicated by (13). The effect of duty cycle variation has been discussed in the previous section.

Stability Requirements

A basic requirement of a Doppler radar system is the generation of a stable RF source for use in the transmitter. Also, since a practical receiver requires an IF greater than zero frequency, the generation of an additional stable radio frequency is required to heterodyne the transmitted frequency to the desired IF.

The frequency stability can be described in terms of allowable modulation on the signal entering the predetection filters of the receiver. For frequency modulation (FM), this is specified in terms of allowable deviation ratio. Short-time stabilities are of basic concern, since long-term stability is not significant as long as the drift does not increase the mutual interference, reduce the power output, or result in an incorrect indication of target velocity. This is in direct contrast to normal frequency-stability specifications where short-term stability is not important, but where long-term stability is.

The permissible spectrum broadening in the receiver is determined by: 1) allowable spurious signals (false targets), 2) loss of sensitivity in presence of interference (clutter), 3) loss of small-signal sensitivity in the interference-free region.

Since the frequency instabilities involved must be

small, it is assumed that the deviation ratio and the fractional modulation are small. The expression for an FM-AM signal modulated by single frequencies is then given approximately as

$$S(\omega) \cong \cos \omega_c t - \frac{\mu}{2} \cos (\omega_c - \omega_a)t + \frac{\mu}{2} \cos (\omega_c + \omega_a)t + \frac{m}{2} \cos (\omega_c + \omega_b)t - \frac{m}{2} \cos (\omega_c - \omega_b)t - \frac{m}{2} \cdot \frac{\mu}{2} \cos (\omega_c - \omega_a + \omega_b)t - \frac{m}{2} \cdot \frac{\mu}{2} \cos (\omega_c - \omega_a - \omega_b)t + \frac{m}{2} \cdot \frac{\mu}{2} \cos (\omega_c + \omega_a + \omega_b)t - \frac{m}{2} \cdot \frac{\mu}{2} \cos (\omega_c + \omega_a - \omega_b)t. \quad (14)$$

Since both m and μ must be small, in order that the sidebands be suitably attenuated, the intermodulation of the FM and AM terms yields negligible coefficients so that the last four terms may be neglected. The remaining terms then are the linear superposition of the AM and FM sidebands. In addition, since the FM modulation consists of a spectrum rather than a discrete frequency, there are all possible intermodulation terms; but again, since the total deviation permitted on the RF frequency must remain small, it is sufficient to merely use superposition.

Generation of Spurious Signals

Spurious signals can appear in the receiver due to modulation of the main-beam and altitude-line clutter signals. In order to assure that these components will not cause spurious signals, it is necessary that the first sideband due to FM at a given modulation frequency be attenuated far below the carrier ($\cong$70 db for typical PD radars). A similar requirement is placed on the AM sidebands due to the same modulating frequency. The final requirement must assure that, during an in-phase relation of the AM and FM sidebands, the composite signal will satisfy the system requirements.

The first sideband amplitude of the FM signal is given by $J_1(\mu)$. $J_1(\mu)$ is approximately $\mu/2$ for $\mu \ll 1$. Now consider a case where all sidebands of 500 cycles per second (cps) or greater must be maintained at 70 db below the carrier. The allowable deviation for the first harmonic term is $\mu_{max} = 6.3 \times 10^{-4}$ for this case. When the modulating frequency is less than 250 cps, a larger number of sidebands can be tolerated. In other words, the allowable deviation increases as the modulating frequency is reduced. Fig. 10 shows the allowable deviation as a function of modulating frequency that is required to maintain a particular sideband below 70 db. This figure is derived from computation of $J_n(\mu)$ and

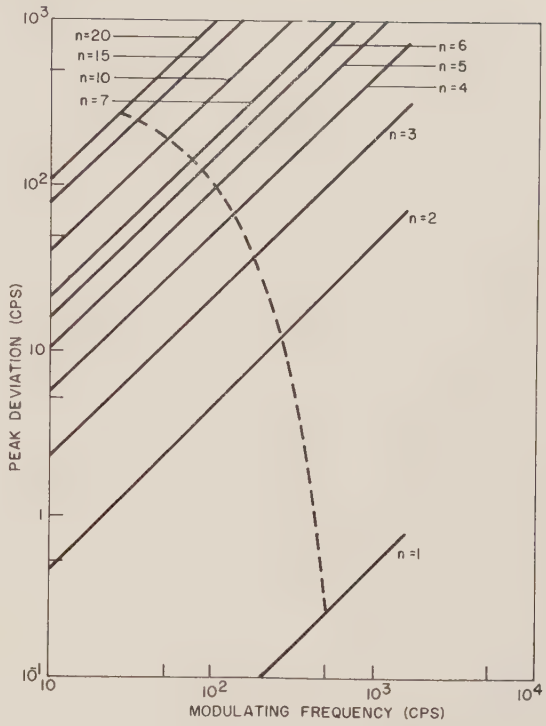


Fig. 10—Maximum deviation allowed to maintain a particular FM sideband about 70 db below the carrier.

has been plotted for various n in terms of the deviation and modulating frequency. The line labeled $n=1$ corresponds to a μ of 6.3×10^{-4} and indicates the $J_1(\mu)$ term.

It is desired in all PD radars that the signal tracking begin X cps from the clutter edge where X is 500 cps for the case assumed here. When the modulating frequency is less than X cps, deviations producing higher-order sidebands can be tolerated. The dashed curve in Fig. 10 defines the maximum peak deviation for a particular modulating frequency such that all sidebands removed 500 cps or more from the carrier will be 70 db or more down from the carrier. For example, the allowable peak deviation for a modulating frequency of 100 cps is shown to be about 100 cps.

For amplitude modulation, only one sideband need be considered. The amplitude of one sideband is $m/2$; hence the allowable value of m is

$$m_{max} = 6.3 \times 10^{-4}.$$

Loss of Signal Energy at the Output of the Predetection Filter

Loss of signal energy occurs in the predetection filter when the signal is frequency modulated. This results from the fact that the sum of the powers in all of the sidebands plus the carrier must be equal to the unmodulated carrier power. When sidebands of the FM signal fall outside of the receiver detection filter pass band, the signal power is reduced. AM will not result in significant signal loss unless the modulating frequency is greater than the filter half-bandwidth.

For the frequency modulation components, three cases must be considered:

1) Small μ and a modulation frequency greater than one half the predetection-filter bandwidth.
2) Large μ and a modulation frequency that approaches zero.
3) A μ that produces significant higher-order modulation components $J_2(\mu)$, $J_3(\mu)$, etc.

For the first case, a typical criterion is that the reduction in $J_0(\mu)$ be 0.1 db. This is equivalent to $J_0(\mu) = 0.9886$ or $\mu = 0.213$.

The second case may be evaluated by assuming that the receiver band-pass filter has a two-pole Butterworth response. This network has a transfer function of the form (where the carrier is much greater than β)

$$\frac{P_o}{P_i} = \frac{1}{1 + \left(\frac{2\Delta}{\beta}\right)^4}. \tag{15}$$

For a large μ as ω_m approaches zero, the power frequency spectrum $H(\omega)$ can be presented as[10]

$$H(\omega) = \frac{1}{\pi\Delta\omega}\left[1 - \left(\frac{\Delta}{\Delta\omega}\right)^2\right]^{-1/2}. \tag{16}$$

Solving for the total power output,

$$\frac{P_o}{P_i} = \frac{1}{\pi\Delta\omega}\int_0^{\Delta\omega} \frac{d\Delta}{\left[1 + \left(\frac{2\Delta}{\beta}\right)^4\right]\left[1 - \left(\frac{\Delta}{\Delta\omega}\right)^2\right]^{1/2}}. \tag{17}$$

Evaluation of this integral gives $\Delta\omega = \beta/4$ for a 0.1-db loss in signal strength. For example, if filters of 500-cycle bandwidth are used, the peak deviation is restricted to 125 cps.

The third case is intermediate between the other two cases. This is the case where $\mu > 0.215$, but is not sufficiently large to satisfy the second case. A diagram can be constructed which satisfactorily specifies all three requirements. The diagram is constructed, using the coordinates of Fig. 10, by drawing a horizontal line from 125-cps deviation until it intercepts the line $\mu = 0.213$.

Requirement on Transmitted Frequency Stability

Let the transmitted carrier be frequency modulated by $\Delta\omega \sin \omega_m t$; then the transmitted frequency ω_T is

$$\omega_T = \omega_c + \Delta\omega \sin \omega_m t. \tag{18}$$

Now assume a point source target moving at a constant radial velocity. The received signal ω_R is shifted by the Doppler shift ω_D. The received signal is

$$\omega_R = \omega_c + \omega_D + \Delta\omega\left(\frac{\omega_c + \omega_D}{\omega_c}\right)\sin \omega_m\left(\frac{\omega_c + \omega_D}{\omega_c}\right)(t - \tau_r)$$
$$\cong \omega_c + \omega_D + \Delta\omega \sin \omega_m(t - \tau_r) \quad \text{for } \omega_D \ll \omega_c. \tag{19}$$

When the transmitted signal is translated by a pure frequency ω_{IF}, the output is[11]

$$\omega_T' = \omega_c - \omega_{IF} + \Delta\omega \sin \omega_m t. \tag{20}$$

When the received signal frequency ω_R is beat against ω_T' in a mixer, the output is

$$\omega_R' = \omega_R - \omega_T'$$
$$= \omega_{IF} + \omega_D + \Delta\omega[\sin \omega_m(t - \tau_r) - \sin \omega_m t]$$
$$= \omega_{IF} + \omega_D + 2\Delta\omega \sin\frac{\omega_m t}{2}\cdot\cos \omega_m\left(t - \frac{\tau_r}{2}\right). \tag{21}$$

From the previous results, the ratio of the frequency deviation in the receiver and transmitter may be found, namely,

$$\frac{f_d'}{f_d} = \left|\frac{2\Delta\omega \sin\frac{\omega_m t}{2}}{\Delta\omega}\right| = 2\left|\sin \pi f_m \tau_r\right|. \tag{22}$$

Examination of this equation shows that for low-modulating frequencies and short ranges, the deviation in the receiver may be considerably less than that of the transmitter; however, for large values of the product $f_m\tau_r$ the deviation in the receiver may exceed the deviation of the transmitter by a factor of two. Fig. 11 shows the curve of $|\sin \pi f_m \tau_r|$ vs f_m for a radar with a maximum range of 130 nautical miles. Curves of $\tau_r < \tau_{max}$ fall to the right of the curve $\tau_r = \tau_{max}$ so that the peaks of the sum of the curves completely fill the area under the envelope curve. The maximum deviation in the receiver can be represented by the envelope. Up to approximately $\pi f_m \tau_r = 0.5$,

$$\left|\sin \pi f_m \tau_r\right| \cong \pi f_m \tau_r. \tag{23}$$

For values of $\pi f_m \tau_{max} > \pi/2$ or $f_m \tau_{rmax} > 0.5$, the envelope of the peak value of $|\sin \pi f_m \tau_{max}|$ can be represented by a constant equal to unity such that $f_d'/f_d = 2$ for $f_m \tau_{r_{max}} > 0.5$.

This envelope can be approximated by two straight lines which intersect at the "corner" given by $f_d'/f_d = 2$ and $f_m = 1/\pi\tau_{max}$ with a slope of -1 to the left of the

[10] S. Goldman, "Frequency Analysis, Modulation and Noise," McGraw-Hill Book Co., Inc., New York, N. Y.; 1948 (p. 154, $1/\pi\Delta\omega$ is a normalizing factor).

[11] This is the common method for developing the local oscillator signal in a PD radar.

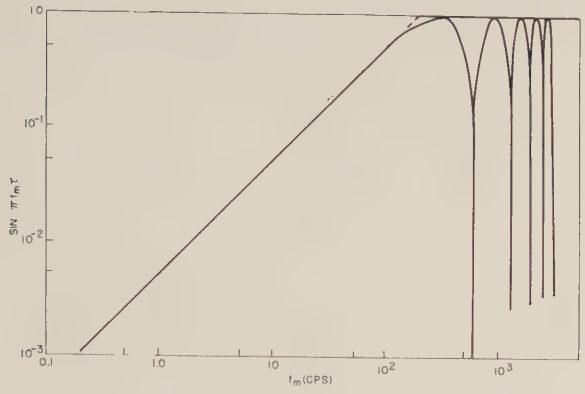


Fig. 11—Frequency deviation of the received signal.

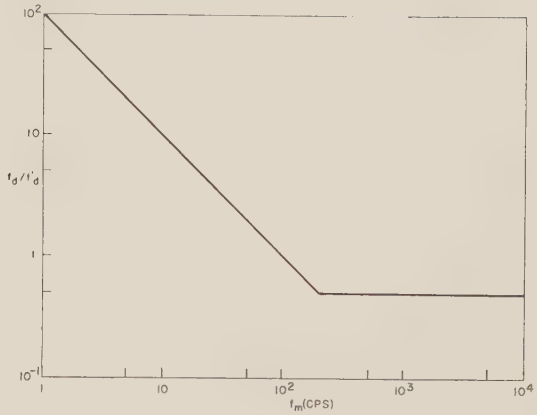


Fig. 12—Frequency deviation of transmitter vs modulating frequency deviation of receiver.

corner. For the previous case of a radar with a maximum range of 130 nautical miles and a permissible frequency deviation of 500 cps in the receiver, the deviation of the transmitter may be plotted. In this case the corner is at $f_d/f_d' = 0.5$ and $f_m = 1/\pi \times 130 \times 12.3 \times 10^{-6} = 200$. This curve is plotted in Fig. 12 and gives the ratio of frequency deviation permissible on the transmitter to that in the receiver.

The above discussion applies when the receiver-beating oscillator is phase locked to the transmitter. However, when two stable oscillators are employed, the restrictions are that at all ranges the worst condition requires that the deviation allowed on the transmitted frequency be only one half the deviation allowed in the receiver IF signal.

Appendix I

Idealized Range Equation

It is shown here that a range-gated pulse-Doppler radar which detects the Doppler shift of the maximum amplitude spectra line has the same range performance as a CW Doppler radar with the same average power.

The idealized range of a CW Doppler radar (S/N = 1) is given by

$$R_0^4 = \frac{P_a G^2 \lambda^2 \sigma}{(4\pi)^3 \overline{NF} K T \Delta F \delta} = A \frac{P_a}{\Delta F} . \tag{24}$$

If a CW wave of amplitude E_p and frequency f_0 is pulsed at a rate f_r with a pulse duration τ, the frequency spectrum $G(f)$ can be expressed as

$$G(f) = E_p \tau f_r \sum_{n=0}^{\infty} U(f_0 - nf_r) \frac{\sin \pi\tau(f_0 - nf_r)}{\pi\tau(f_0 - nf_r)} \tag{25}$$

where E_p is the amplitude of the sine wave f_0.

The maximum amplitude spectra line is at f_0, or

$$G(f_0) = E_p \tau f_r. \tag{26}$$

The duty cycle d is equal to $f_r\tau$. For this spectral line, the average power is $(E_p^2 d^2)/2$.

Substituting in the range equation (1),

$$R_0^4 = A \frac{E_p^2 d^2}{2\Delta F} . \tag{27}$$

Recognizing that the peak power P_t is $E_p^2/2$ and the average power P_a is equal to $P_t d$, the equation becomes

$$R_0^4 = \frac{P_a d}{\Delta F} . \tag{28}$$

If the system is range gated with a pulse equal to the transmitted pulse width, then the noise ($\overline{NF}kT\Delta F$) is reduced by d, or

$$R_0^4 = A \frac{P_a}{\Delta F}$$

which is the same as (1).

If the range-gate width is greater than τ, then the output noise is increased. If the duty cycle of the noise is represented by d_n, the range equation can be expressed as

$$R_0^4 = A \frac{P_a}{\Delta F} \cdot \frac{d}{d_n} . \tag{29}$$

The Effect of Changing Frequency and Antenna Size

The idealized range R_0 can be written

$$R_0 = \left[B \frac{P_a G^2 \lambda^2}{\Delta F} \right]^{1/4} . \tag{30}$$

Assume that a fixed solid angle of space ϕ must be searched in a frame time T_f, then

$$T_f = \frac{\phi}{\theta^2} t_d. \tag{31}$$

For a fixed sample time

$$\Delta F = \frac{1}{t_d} . \tag{32}$$

Substituting in (31) and solving for ΔF,

$$\Delta F = \frac{\phi}{T_f} \cdot \frac{1}{\theta^2} .$$

Noting that θ^2 is proportional to $1/G$, (30) under these conditions becomes

$$R_0^4 = B' P_a \lambda^2 G. \tag{33}$$

Now

$$G \cong \pi^2 \frac{D^2}{\lambda^2} \tag{34}$$

where D is the diameter of the antenna.

Substituting in (33),

$$R_0 \propto \sqrt{D \sqrt{Pa}}. \tag{35}$$

It is observed that for a fixed antenna size and under the stated conditions for a PD radar R_0 is independent of frequency.

Symbols

A = Constant.
AM = Amplitude modulation.
B = Constant.
β = Butterworth filter bandwidth in radian frequency.
C = Constant.
CW = Continuous wave.
cps = Cycles per second.
D = Parabolic antenna diameter.
d = Duty cycle.
d_n = Noise duty cycle.
Δ = Detuning from the center frequency of a filter.
E_p = Transmitted sine wave amplitude.
δ = System losses.
FM = Frequency modulation.
F_v = Total Doppler frequency range searched by radar.
ΔF = Effective receiver noise bandwidth.
f_{IF} = Receiver intermediate frequency.
f_d = Frequency deviation in the transmitter.
f_d' = Frequency deviation in the receiver.
f_{d_A} = Doppler frequency corresponding to maximum ground velocity of the interceptor aircraft.
f_{d_T} = Doppler frequency corresponding to maximum ground velocity of target.
f_{LO} = Local oscillator frequency.
Δf = Half-power filter bandwidth.
f_m = Modulation frequency.
f_0 = Carrier frequency.
f_r = Pulse recurrence frequency (PRF).
G = Antenna gain.
θ = Antenna half-power beamwidth, one way.
IF = Intermediate frequency.
$J_n(X)$ = Bessel function of the first kind and nth order.
k = Boltzmann's constant.
λ = Operating wavelength.
μ = Frequency deviation ratio $\Delta\omega/\omega_a$.
m = Fractional amplitude modulation.
N = Number of signal variants integrated.
$n = T_{fa} \cdot \Delta F$ = Number of opportunities for noise to exceed the threshold in one false alarm time.
$\overline{NF}$ = Receiver noise figure.
P = Range and velocity search intervals.
P_a = Average transmitted power.
P_c = Cumulative probability of detection.
PD = Pulse Doppler.
P_{dn} = Probability of detection on the nth look.
P_i = Filter-power input.
P_o = Filter-power output.
PRF = Pulse recurrence frequency.
P_t = Peak transmitter power.
P_{tc} = Total cumulative probability.
ϕ = Solid angle search volume.
ΔR = Target distance closed in a frame time.
RF = Radio frequency.
R_0 = Idealized radar range, signal-to-noise ratio = 1.
R_{0_e} = Effective idealized range for eclipsed signal.
S/N = Signal-to-noise ratio.
σ = Target size.
$S(\omega)$ = Frequency spectrum in radians per second.
T = Absolute temperature.
t_d = Antenna dwell time on target.
T_{fa} = False alarm time.
T_r = Pulse period (= 1/PRF).
T_t = Target search time.
τ = Transmitter pulse width.
τ_r = Radiation round trip time.
ω = Radian frequency.
ω_a = FM modulation frequency in radians per second.
ω_b = AM modulation frequency in radians per second.
ω_c = RF carrier frequency in radians per second.
ω_D = Doppler shift in radians per second.
$\Delta\omega$ = Peak deviation of the carrier in radians per second.
ω_{IF} = Transulation frequency in radians per second.
ω_m = Modulating radian frequency.
ω_{max} = Maximum antenna angular velocity.
ω_R = Received signal in radians per second.
ω_T = Transmitted frequency in radians per second.

Acknowledgment

Many members of the Westinghouse Air Arm Engineering staff contributed to the concepts and developments given in this paper. Among the principal contributors are M. Dempsey, W. Ewanus, H. Grauling, D. Healey, H. B. Smith, W. Skillman, and D. Mooney.

A High-Resolution Radar Combat-Surveillance System*

L. J. CUTRONA†, SENIOR MEMBER, IRE, W. E. VIVIAN‡, E. N. LEITH†, AND G. O. HALL†, MEMBER, IRE

Summary—**An account is given of the development of the AN/UPD-1 (XPM-1) system. This airborne mapping radar, by synthesizing an extremely long antenna which expands in length in direct proportion to radar range, provides a linear resolution in the azimuth direction that is constant for all radar ranges.**

I. Introduction

ON April 19, 1960 at the Washington, D. C., National Airport, the U. S. Army unveiled a new type of combat-surveillance radar system which is designated as the AN/UPD-1 (XPM-1). The principal elements of the system are an airborne side-looking coherent radar and a ground van housing radar-data processing equipment. The unique feature of the system is the data processing that produces the equivalent effect of an extremely long antenna which expands in length in direct proportion to radar range. This, in principle, can provide a linear resolution in the azimuth direction that is constant for all radar ranges.

This combat-surveillance system represents a dividend from a continuing research program in radar at the Institute of Science and Technology of The University of Michigan. This radar research is a part of Project MICHIGAN, which is sponsored by the U. S. Army Combat Surveillance Agency of the U. S. Army Signal Corps. Four models of the AN/UPD-1 (XPM-1) experimental systems were produced through separate funding for Project MICHIGAN. The airborne portion of the system was subcontracted to Texas Instruments, Inc., of Dallas, who in turn subcontracted to General Precision Laboratories and Kearfott, Inc., for Doppler inertial subsystems. The ground vans containing the radar-data handling and processing equipment were fabricated by the Institute of Science and Technology of The University of Michigan.

The AN/UPD-1 system is a good example of how a team effort of a university and industry can provide experimental models of new devices for early tactical evaluation. In less than ten months from the first successful demonstration with a breadboard system, a subcontract was in force and work was under way on the construction of the experimental systems.

It is the purpose of this article to trace the development of this system from its inception, to discuss the basic concepts of the system and its components, and to present some of the high-resolution radar photographs obtained with it.

II. The Fundamental Idea

In the summer of 1953, scientists from all parts of the country met at the University with representatives of the U. S. Army and personnel from The University of Michigan. This summer study, known as Project WOLVERINE, was for the purpose of recommending research and development programs leading to an improved capability for securing information about the battle area.

The basic idea behind the high-resolution radar was advanced in a subgroup considering the application of radar to combat surveillance. The group recognized that the major shortcoming of radars was in angular resolution. In combat surveillance the objects of military interest are against a background of the surrounding terrain; therefore, unless a radar has fine resolution in both range and azimuth, the radar return from a target of military interest is likely to go undetected because of the large clutter return from the background. With conventional techniques, fine resolution in range was readily attainable with short-duration radar pulses. Since with conventional techniques angular resolution is directly related to the radar antenna beamwidth, the only avenues for improvements in azimuth resolution were through an increase in antenna size, an increase in operating frequency, or a combination of both. Although the largest feasible size for an airborne antenna and the highest practical radar frequency (dictated by available peak power sources and acceptable attenuation for various weather conditions) were to be used, it was evident that the angular resolution was inadequate for the intended surveillance functions at radar ranges beyond a few miles.

A new idea for producing the equivalent effect of an extremely long antenna became the subject of discussion by Dr. C. Sherwin of the University of Illinois, W. Hausz of the General Electric Company, J. Koehler of the Philco Corporation and Dr. Louis Cutrona of the University of Michigan. It was thought that an extremely long antenna could be synthesized by using the forward motion of the aircraft to carry a side-looking radar antenna to positions which could then be treated as though they were the individual antenna elements of a linear antenna array.

The fundamental idea can best be grasped by consideration of Fig. 1. For the physical antenna array shown in Fig. 1(a), the individual transmitting and receiving elements are dipoles. While each dipole has a

* Received by the PGMIL February 21, 1961. This work was conducted by Project MICHIGAN, under Department of the Army Contract DA-36-039 SC-78801, administered by the U. S. Army Signal Corps.

† The Inst. of Sci. and Tech., The University of Michigan, Ann Arbor, Mich.

‡ Conductron Corp., Ann Arbor, Mich.

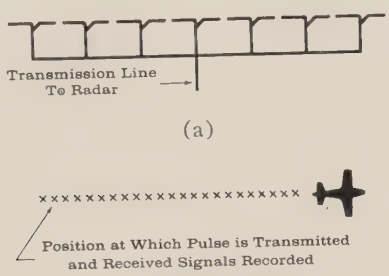


Fig. 1—(a) Physical array of dipoles. (b) Synthetic-array generation.

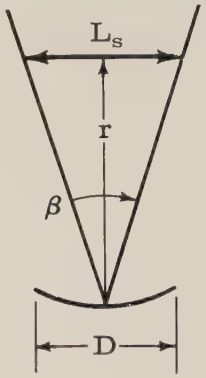


Fig. 2—Synthetic-antenna geometry.

broad radiation pattern, the assemblage of dipoles is made to produce a narrow antenna beam by making the electrical lengths of the individual transmission lines such that signals arriving in phase at the dipoles are added in phase at the main transmission line to the radar. Since transmission and reception at each of the dipoles is simultaneous, this fixed adjustment of phases serves to maintain the desired beam pattern.

Fig. 1(b) depicts the generation of the synthetic array. In this case the individual elements of the array, as indicated by the x's, do not exist simultaneously. Starting with the position at the extreme left, a radar pulse is transmitted and the return signals recorded. A short time later the aircraft has carried the side-looking antenna to the second position where another pulse is transmitted and the received signals recorded. In this way a signal is recorded for each of the positions of the synthetic array. To achieve the effect of a linear array such as Fig. 1(a), the return signals must contain phase information which is preserved in the recording. Then by appropriate data processing it should be possible to retrieve the stored data and combine it in proper phase to synthesize the desired effect of a narrow-beamed antenna.

It was evident that the radar would have to have excellent transmitter frequency stability and a stable frequency reference for use in comparing the phase of the return signals with the phase of the transmitted pulses. The method of signal storage (recording) had to preserve the range information so that the data processing could be accomplished separately for each element of radar range.

The well-known formula for antenna beamwidth showed that, at least in theory, the synthetic-antenna concept had a great potential for fine angular resolution. The half-power beamwidth β of the physical side-looking antenna (Fig. 2) is

$$\beta = k\frac{\lambda}{D} \text{ radians} \tag{1}$$

in which

λ is the wavelength of the radar,
D is the length of the physical aperture, and
k is the illumination factor (greater than unity).

The distance across the antenna beam for a radar range r is βr, which represents the amount of data that can be collected between half-power points for this range. If this amount of data is processed at each range, this would represent a synthetic antenna length

$$L_s = \beta r \text{ feet} \tag{2}$$

if r is expressed in feet. The phase information obtained for each of the element positions of the synthesized antenna is based on the round-trip distance between the side-looking antenna and echoing objects; for the physical array the phase depends on the one-way path length. As a consequence of this difference, the phase difference between elements of equally spaced elements of a synthetic array is twice that of a physical array with the same spacing; a synthetic antenna has half the beamwidth of a physical antenna of a given length. That is,

$$\beta_s = k_1\frac{\lambda}{2L_s} \text{ radians} \tag{3}$$

in which k_1 is the illumination function for the synthesized antenna. Using (1) this can be expressed in terms of the size of the physical antenna and radar range as

$$\beta_s = \frac{k_1}{k}\frac{D}{2r} \text{ radians.} \tag{4}$$

This indicates that the synthetic beamwidth β_s is not only independent of frequency, but also decreases with radar range.

To obtain a measure of resolution, the usual assumption was made that targets could be resolved in angle if they were separated by one antenna beamwidth. For the beamwidth β_s of the synthetic antenna, the distance across the beam at any radar range r is

$$\Delta X = \beta_s r = \frac{k_1}{k}\frac{D}{2} \text{ feet} \tag{5}$$

if the length of the physical antenna aperture D is expressed in feet.

In theory, then, the resolution in the azimuth direction for a synthetic antenna radar is independent of range, independent of radar frequency, and smaller than the physical length of the actual side-looking antenna carried by the aircraft. This assumes the generation of a synthetic antenna of a length equal to the distance across the radar beam at each radar range.

III. The Research and Development Program

When Project MICHIGAN was inaugurated in the fall of 1953, a small group undertook an evaluation of this new concept for radars. Analytic expressions were obtained for the received data and the theoretical azimuth resolution was verified with filter theory. Expressions were also obtained for the operations to be performed by the data processor. Because of the achievable lengths of synthetic antennas, it was not feasible to neglect the curvature of the phase front of received signals. It would be impossible to approach the theoretical angular resolution unless the antenna were focused. Since the curvature of the phase front varies with radar range, the data processor needed to be one which applied a focusing function that varied with range.

As this analysis progressed, the group became convinced that the synthetic-antenna technique held promise and that a breadboard system should be built for experimental evaluation. Photographic film was selected as the storage medium for the received radar signals. With film as a storage medium, three dimensions were available for recording: distance across the film, distance along the film, and the exposure. Film had the further advantage of providing a permanent record which could be processed many times using different processing parameters. Digital processing techniques for generating the synthetic antenna did not appear promising because of the number of bits of information to be handled and the need for focusing functions which differed for each segment of radar range. After a period of basic laboratory experimentation an analog data-processing technique was evolved which was not only simple, but which also provided simultaneous focusing of the radar data at all radar ranges.

By the end of 1954, a great many of the components for the radar system had been proven by laboratory experimentation and microwave components for the radar were on order. It was decided to build a preliminary model of the coherent radar in the laboratory and to check its performance with a roof-mounted antenna before proceeding with the fabrication of an airborne breadboard. In this way enough experience could be gained to finalize a design before an aircraft became available. A stabilizing cavity and a crystal-controlled reference at the IF frequency were selected as the means for obtaining the necessary radar frequency stability and coherence. With this scheme, the prime frequency reference is a continuously operating klystron stabilized by a high Q cavity. The second reference at the IF frequency, also continuously operating, is used to single-sideband modulate the RF frequency. A chain of pulsed klystron amplifiers is then used to produce the desired RF pulse shape and to boost the power level for transmission. On reception, the prime frequency reference serves as the local oscillator for heterodyning the received signals.

The laboratory breadboard was operating with a roof-mounted antenna by the fall of 1955. Standard laboratory oscilloscopes were used as an output display and a modified scope camera was used for recording. In the spring of 1956 work was started on the fabrication of the components for an airborne breadboard system.

A C-46F aircraft was obtained on bailment from the U. S. Air Force in the fall of 1956. Aircraft overhaul, modification, and radar equipment installation were completed in July of 1957. Figs. 3 and 4 show external and internal views of the aircraft with its installed equipment. The radome houses the side-looking radar antenna and the antenna for Doppler navigation radar.

Fig. 3—The C-46F.

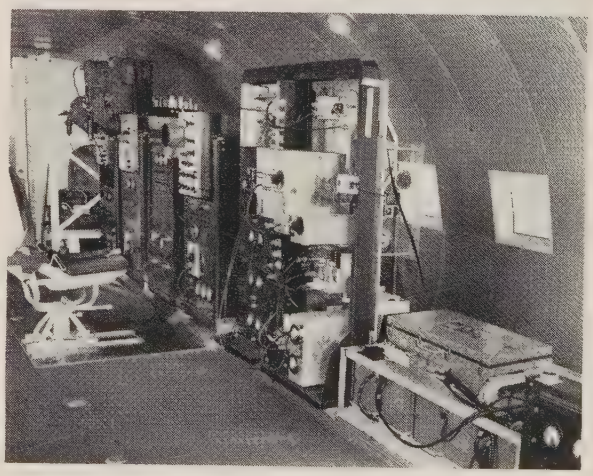


Fig. 4—Breadboard system in the C-46F.

In the foreground of the interior view is a 14-channel Ampex tape recorder which is used to obtain records of aircraft accelerations, antenna orientation, and various voltages within the radar system. Tape records were extremely valuable for post-flight evaluation, for trouble-shooting equipment, for formulating modification requirements, and for improving flight-operation procedures. The first rack of the radar proper houses the coherent transmitter, which is followed in turn by the

main control and operators' panel, power supplies, the receiver, and two radar signal displays with camera recorders. The components for the Doppler navigator are mounted in a separate frame at the extreme forward position in the cabin.

For this breadboard system, standard laboratory Tektronix 535 oscilloscopes with P-11 phosphor cathode-ray tubes were used. The recording cameras, which were greatly modified Fairchild scope cameras, continuously advanced 35-mm film in a direction perpendicular to an intensity-modulated range trace on the display. The rate of advance of the film was controlled by ground-speed information obtained from the Doppler navigator. The repetition rate of the radar was also ground-speed controlled to maintain constant film exposure. The radar antenna was maintained perpendicular to the ground track through a combination of drift-angle information from the Doppler navigator and a

Fig. 5—The AN/UPD-1 system.

Fig. 6—Interior of the L-23D.

Fig. 7—Data processor.

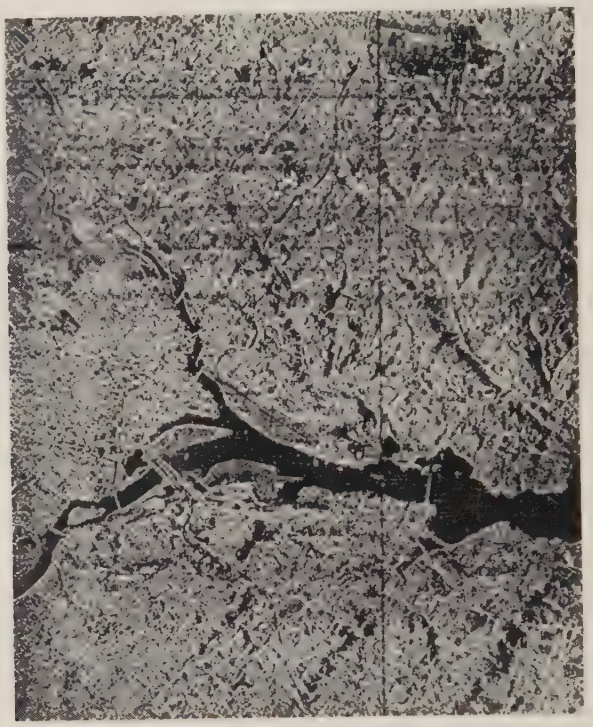


Fig. 8—AN/UPD-1 map of Washington.

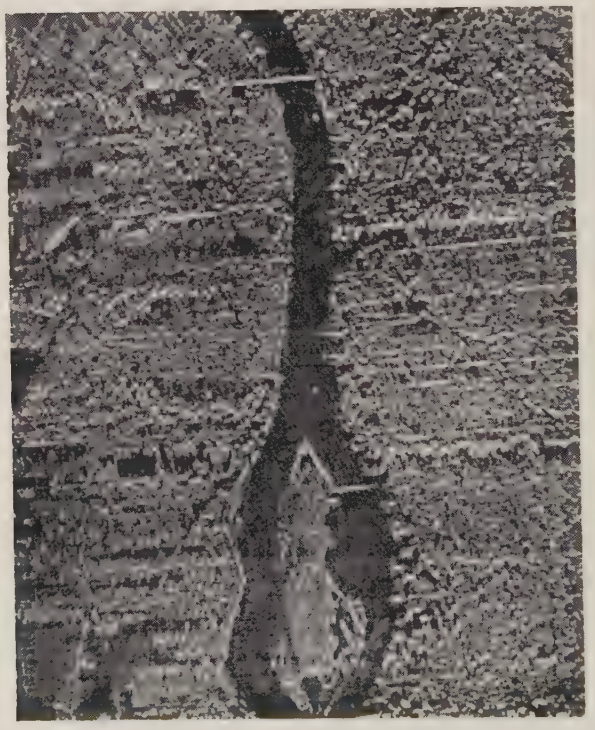


Fig. 9—AN/UPD-1 map of Detroit.

gyro mounted on the antenna base. The navigator established the long-time average while the gyro corrected for short-time yawing.

Success was achieved on August 23, 1957. This was the fifth flight on which data were recorded. After this one successful flight, which demonstrated high angular resolution with the synthetic-antenna technique, it was two months before another successful flight was achieved. As with most R&D programs there was a period in which everything seemed to go wrong. However, there were some results which were encouraging.

Equipment reliability was improved and a number of fruitful flights were made in the Willow Run area. In the spring of 1958 high-resolution strip maps were obtained of the Fort Huachuca area and of Tucson, Arizona.

By this time, the capability of the synthetic-antenna technique was recognized and steps were taken to initiate the procurement of experimental models mounted in U. S. Army aircraft.

The research and development program with the C-46 "flying radar laboratory" is being carried on and the results of the experimentation will be made available for the next generation of high-resolution radars.

IV. The AN/UPD-1 (XPM-1) Systems

In 1958, at the request of the U. S. Army Combat Surveillance Agency, specifications were prepared and a subcontract let to Texas Instruments, Inc., for four experimental models of the airborne portion of the high-resolution radar system installed in L-23D aircraft. The ground portion of the system was fabricated at Willow Run Laboratories and installed in U. S. Army expansible vans. Work started on July 1, 1958.

The first of the four models of the airborne equipment arrived at Willow Run Laboratories in the fall of 1959. Successful high-resolution operation was achieved in the Dallas area, in Michigan, and in the area around Washington, D. C. On April 19 a press conference was held in Washington and some high-resolution radar photographs were cleared for release to the press. Two complete systems were on display at the Washington National Airport.

Fig. 5 shows the aircraft and the ground processing van. As in the C-46F aircraft, the radome houses both the side-looking radar antenna and a Doppler navigator antenna. The van houses, in addition to film-development equipment and a data processor, an enlarging viewer to display high-resolution strip-map photographs at standard map scale. Fig. 6 shows the equipment in the L-23D. The complete weight of the radar including Doppler-inertial auxiliaries is 700 lbs. Fig. 7 is a photograph of the data processor in the processing van.

Figs. 8 and 9 are high-resolution radar photographs made with the AN/UPD-1 systems.

The Evolution and Application of Coherent Radar Systems*

N. R. GILLESPIE†, senior member, ire, J. B. HIGLEY†, and N. MacKINNON†

Summary—**A brief introduction to coherent radar applications is given by discussing some basic pulse and CW systems. Comments on the historical development of these systems are followed by a general discussion of coherent radar parameter variations as they relate to the accuracy, resolution and ambiguity of target position or speed measurements.**

Introduction

THE USE of coherent radar systems has been greatly accelerated in recent years. Military and civilian applications such as Doppler navigation, ballistic missile tracking, speed measuring and airways surveillance have been the inspiration of many specific papers on coherent systems.

The purpose of this paper is to provide a uniform coverage of this subject, including some history, a discussion of limitations and examples of the wide variety of coherent radar systems.

A coherent system utilizes the information contained in the phase relation of a received signal with respect to the transmitted signal. The noncoherent types of pulse radar cannot take advantage of this phase information, and therefore experience difficulty if the target return is buried in reflections from nearby land or water.

The ability of a coherent system to measure the relative phase of target returns enables the Doppler frequency shift, caused by the relative target motion, to be used for multiple return discrimination.

For example, the moving target indicator (MTI) search radar will display targets only if the Doppler

* Received by the PGMIL, January 17, 1961.
† Raytheon Co., Wayland, Mass.

shift is sufficiently greater than zero and different from multiples of its pulse repetition frequency. Modern CW missile guidance radar intelligence circuits are satisfied with nothing less than a Doppler return within a specified narrow bandwidth set in advance, or determined by a Doppler tracking speed gate.

CW systems are often used when velocity information is more important than range. The coherent pulsed system, on the other hand, is usually more concerned with target range than its velocity with respect to the system. As pulse widths increase and CW systems become modulated, the number of similar characteristics increases and the two systems tend to coalesce.

The Evolution of Coherent Radar

In 1922 Dr. A. Hoyt Taylor and Leo C. Young observed that the passage of a moving steamer caused distortion of their received radio signals.[1] This distortion, similar to that experienced by TV viewers living near airports, was the "beating" or intermodulation of the directly received CW transmission with the steamer-reflected Doppler-shifted version of that same transmission.

The subsequent development of CW "radio echo" moving target locators was undertaken by both the U. S. Navy and U. S. Army in the 1930's. Other countries were working on this potentially useful technique, and by 1935 the Germans were experimenting with 3000-Mc CW moving target detection devices. Some of the detection devices appeared on tugboats and were intended for harbor guidance.[2]

During World War II, operators of pulsed search radars noticed that the returns from moving targets, which were at the same range as stronger ground reflections, could sometimes be detected on an ordinary "A" display (signal amplitude vs range) by a fluttering of the stronger "clutter" return at the moving target range. This phenomena, often called the "butterfly" effect, could be noticed only when the clutter return did not saturate the radar receiver. The subsequent invention of the lin-log IF amplifier response to increase receiver dynamic range advanced the utilization of the butterfly moving target detection technique. The resultant system came to be known as the noncoherent MTI, because the transmitted signal phase reference necessary for detecting targets in clutter was held not by a local or internal "memory," but by external and fortuitous means, namely uncontrolled clutter. Systems of this type were sometimes used to monitor road intersection activity under blackout battlefield conditions. Fluttering of the ground return at the right range was the signal for previously aimed shell fire.

Since clutter does not necessarily occur at all ranges, some internal means was necessary to differentiate reliably between a water tower and an attacking aircraft. The coherent MTI was born with the addition of the internal phase reference.

Several means of remembering or controlling the transmitted phase characteristic were possible. A more obvious technique is shown in Fig. 1. During the war, when the early coherent MTI development was started, the required stable master oscillator (STAMO) could be designed, but pulsed microwave RF power amplifiers were not yet within the state of the art. Consequently, this coherent technique was not used until recently. The pulsed magnetron was the only practical means of generating high transmitted power in microwave regions where the enemy had not yet learned to jam.[3]

Since the RF pulses emitted from a pulsed magnetron do not bear any predictable phase relationship to each other, a free-running oscillator could not accurately store the transmitted phase information, and another approach was necessary. The concept of phase pulling and locking an oscillator to each transmitted pulse as it was generated was exploited. This could be accomplished at RF or IF, but the method which was most widely accepted is shown in Fig. 2.[4] The stable local oscillator (STALO) needed stability of a small fraction of a cycle for the short time of one pulse repetition period. The phase-locked coherent oscillator (COHO) had to have a Q low enough to allow rapid phase pull-in. With this technique the transmitted phase and frequency were remembered for pulse reception and substantially the same performance as the system in Fig. 1 was realized.

In terms of the noncoherent MTI, the coherent system supplied its own clutter reference, and targets could be detected by observing the butterfly effect on an "A" display. However, this type of presentation was not as effective as the PPI for searching in azimuth, and techniques were needed for displaying only moving targets which could utilize the phosphor storage and map-like presentation of the PPI display.

Because successive video returns from stationary targets are similar, and those from the moving targets flutter up and down at the Doppler frequency, a clutter cancellation scheme could be devised. Fig. 3 illustrates the one-delay-line subtraction technique. It can be seen intuitively that, when the video time delay exactly equals the transmitted pulse spacing, similarities in successive returns will be cancelled. Clutter rejection schemes are discussed in more detail in later paragraphs.

[1] "Radar, A Report on Science at War," U. S. Govt. Printing Office, Washington, D. C.; 1945.
[2] "Microwaves to detect aircraft," *Electronics*, vol. 8, pp. 18–19; September, 1935.
[3] J. P. Baxter III, "Scientists Against Time," Little, Brown and Co., Boston, Mass.; 1946.
[4] L. N. Ridenour, "Radar System Engineering," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1, ch. 16; 1947.

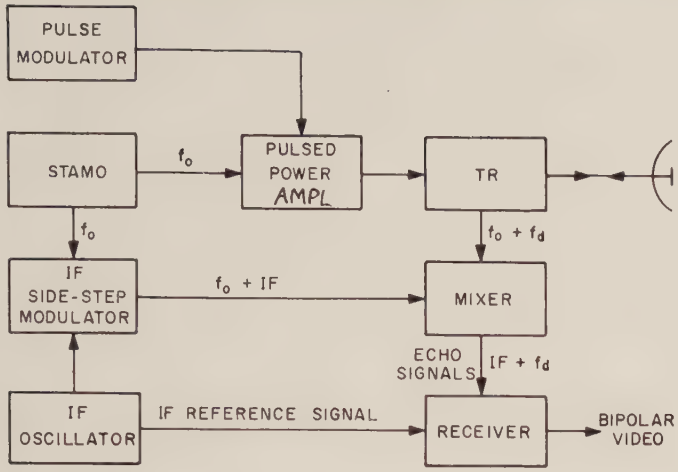


Fig. 1—A fully coherent pulsed radar system.

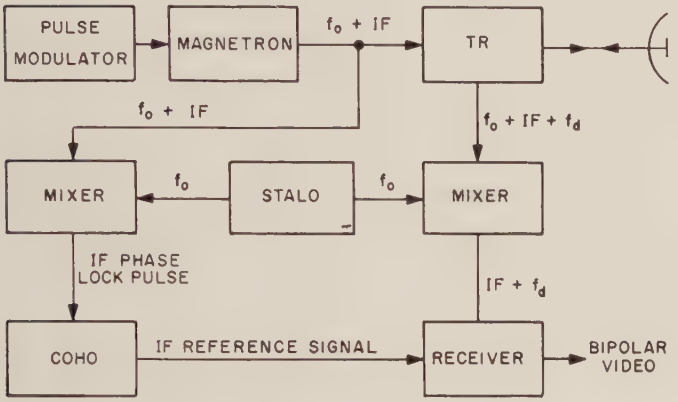


Fig. 2—A phase-locked coherent MTI.

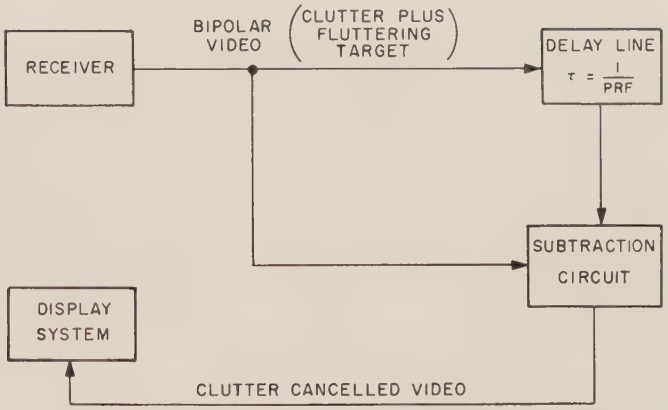


Fig. 3—One-delay subtraction clutter canceller.

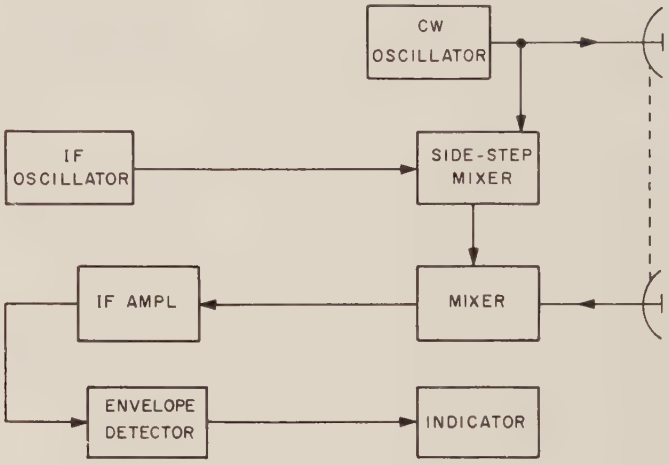


Fig. 4—A simplified CW radar system.

The coherent MTI radar described in the past few paragraphs did not get into field usage during World War II, but it has been used extensively since then for airways and airport surveillance as well as for military applications.

CW systems similar to those originally developed in the 1930's for moving target detection were further refined so that some measure of target range could be taken. Two advantages that CW radar enjoyed over ordinary pulsed radar were the ability of the system to detect targets at very short ranges, where the pulse radar receivers were still desensitized by the finite transmit-receive duplexer recovery process, and the ability to detect low-flying aircraft or moving vehicles, where the reflections from the nearby ground would have reduced normal pulse radar effectiveness.

A simplified CW radar system is shown in Fig. 4. The Doppler shift caused by a moving target is obtained by comparison of the direct and reflected received signals in the receiver. A target indicator for this system may be a pair of earphones, which, together with an operator's sense of hearing, can provide a very sensitive target detector whenever the maximum expected Doppler shift is less than a few thousand cycles. Systems of this type were used for aircraft detection and measurements of shell velocities. A modified version of this CW system is the speed-measuring police radar, in which a frequency-reading meter calibrated in velocity replaces the human ear.

Power considerations aside, reasonable determination of range with a CW system is not possible without "time-tagging" the transmitted energy in some fashion. AM and FM have been used for this purpose. A specialized form of AM, *e.g.*, pulse modulation, could result in a system identical to the coherent MTI in Fig. 1. The most common form of FM CW is the linearly frequency swept transmission of the radio altimeter. Ranges down to within a hundred feet or so are desired, and the Doppler shift due to vertical motion of the aircraft or changing terrain serves only to introduce range errors (the amount being dependent on the system parameters). Altitude is determined by measurement of the frequency difference between the reflected signal and the transmitted signal at any given time. Since the transmitted frequency is changing uniformly, the measured difference may be calibrated in terms of the echo transit time.

Signal Processing Techniques

Fig. 5 illustrates some techniques which, along with the delay line cancellation scheme, have the most common usage in coherent radars. Fig. 5(a) is an example of a range-gated, clutter-filter technique which might be used in an automatic detection system when the system range is covered by the use of range gates. This system (with some modification) might also be used in an automatic range tracking scheme or in a range and Doppler frequency measuring device. It demonstrates the use of clutter filters with a range-gated system and the use of contiguous filters. The lower cutoff in the clutter filter is chosen to reject clutter movements up to some specified rate which might be caused by the motion of waves on the surface of the sea or windblown chaff. The upper cutoff frequency is set at PRF/2, for the example given, in order to pass only one spectral line of the return signal. There are many variations of these techniques, including ones which operate at an intermediate frequency, which are beyond the scope of this paper.

The diagram in Fig. 5(b) shows a method of measuring Doppler frequencies by use of a closed-loop speed gate. The speed gate will act both as a filter with a narrow band-pass centered on the Doppler return and as a device for Doppler frequency measurement. The speed gate and the contiguous filter set can be used in either CW or pulse systems, depending on the particular application.

The periodic frequency response of the clutter filter shown in Fig. 6(a) is approached by the delay line subtraction technique discussed previously in conjunction with Fig. 3. The frequency response for this conventional MTI clutter canceller is shown by Fig. 6(b) as the absolute value of a sine wave.[4] With this response, only nonmodulated clutter returns can be completely cancelled, and partial cancellation will occur for desired target returns at frequencies other than one-half the PRF.

By cascading two of these simple cancellation circuits, the sine-squared response shown in Fig. 6(c) can result. With this circuitry, clutter at frequencies near zero, or a multiple of the PRF, is attenuated more than is possible with the one-delay-line circuit. Through the use of recirculation or feedback techniques, the response shown in Fig. 6(a) may be approached.[5]

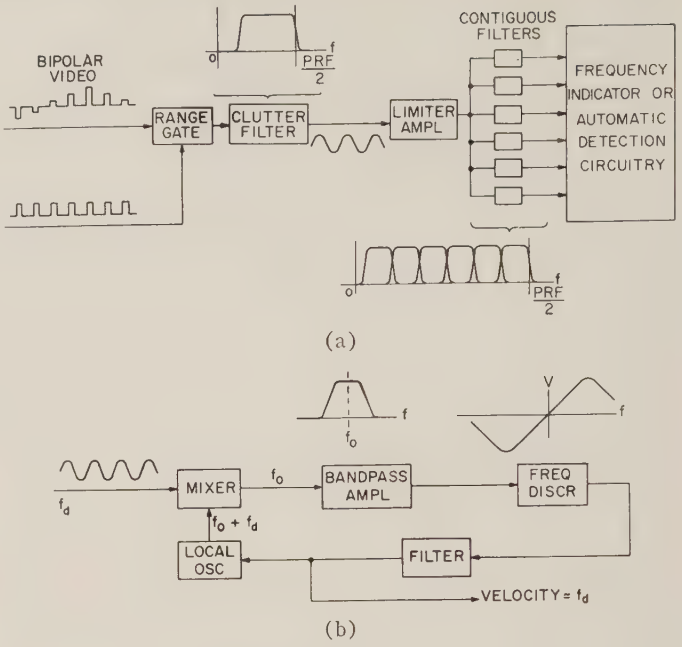


Fig. 5—Signal processing techniques. (a) Coherent pulse receiver; (b) CW Doppler speed gate.

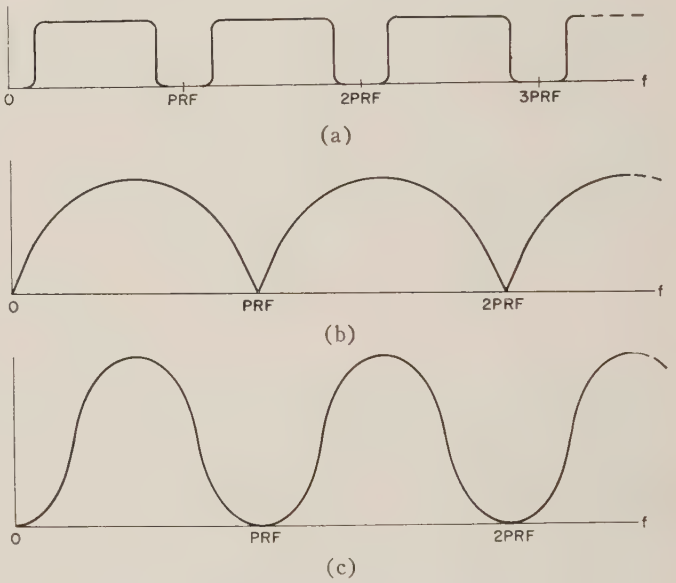


Fig. 6—Velocity response characteristics. (a) Frequency response for coherent pulse system with clutter filter. (b) Frequency response for a one-delay-line subtraction clutter canceller. (c) Frequency response for a two-delay-line subtraction clutter canceller.

Coherent System Spectra

Since system analysis of coherent radars usually involves an examination of signal spectra, this section is devoted to the consideration of spectra and waveforms. The two extremes of coherent radars may be considered to be the unmodulated CW system, with a single-frequency component which is continuous in time, and the impulse radar with zero time duration and a continuous frequency spectrum.

The unmodulated CW signal contains unambiguous target speed information in the Doppler shift and highly ambiguous range information in the relative phase. The target range information repeats itself or is ambiguous every half wavelength, making this information generally useless at microwave frequencies. Multiple targets may be resolved on the basis of frequency if their speed difference is adequate, but when they are at different ranges and the same Doppler frequency they cannot be resolved.

[5] "Advanced Radar Design Techniques," Equipment Div., Raytheon Co., Wayland, Mass., Sec. 1; September, 1960.

The impulse radar, on the other hand, yields target range information as a time delay, with sharp resolution of multiple targets in range. The impulse spectrum, however, having included all frequencies, does not allow frequency discrimination and therefore has no means of resolving multiple targets at the same range.

The above examples are obviously extremes. There are a wide variety of radars between these extremes, and they may be grouped into two general categories: modulated CW and pulsed Doppler radar.

Fig. 7 shows the time and frequency diagrams for a coherent pulse radar signal.

If there were no phase continuity between pulses in the train, the energy in the spectral lines of Fig. 7(b) would be spread continuously under the spectrum envelope. A COHO receiver system can remember the phase of each pulse and unscramble the phases in the spectrum, thus restoring the spectral lines in the receiver to those which would have been received with fully coherent transmission.

With a limited observation time, such as caused by antenna scanning, the frequency spectrum no longer consists of the familiar lines of a pulse train with unlimited observation time. Instead, the spectral lines are modified by the frequency transform of the observation time characteristic. If the observation interval is inspected periodically, as in a scanning radar, this modification will have a fine line structure with a spacing equal to the reciprocal of the time duration between observation intervals. If the observation time is reduced to include only one pulse, the frequency spectrum will contain all frequency components instead of discrete lines, as one would expect from the Fourier transform of a single pulse. It will be shown later that observation time becomes an important consideration in the study of accuracy and resolution.

Fig. 8 shows the relationship between transmitted and received signals in time and frequency (with the effect of finite observation time ignored), where the transmitted signals are shown in solid lines and the received signals in dashed lines. The signals from this pulsed Doppler radar are shifted in time $\Delta t = 2R/c$, where R = range to target and c = velocity of propagation.

The signals are shifted in frequency $\Delta f = (c+v/(c-v)f$, or approximately $2vf/c$, where v = target velocity and f = frequency of particular spectral component.

Ambiguities or uncertainties in range will exist when $R \geqq c/2f_r$, where f_r = pulse repetition rate, and ambiguities in frequencies will result when $\Delta f \geqq f_r$ 2.

A Doppler shift Δf, large enough to produce ambiguous velocities, is shown in Fig. 8(b). The closest line-to-line spacing is measured as $\Delta f'$. There are many techniques for solving ambiguities by varying one or more radar parameters. The principles are discussed later in this paper.

A pulsed Doppler radar is essentially a range-measuring CW radar with 100 per cent amplitude modulation by a rectangular pulse. As previously mentioned,

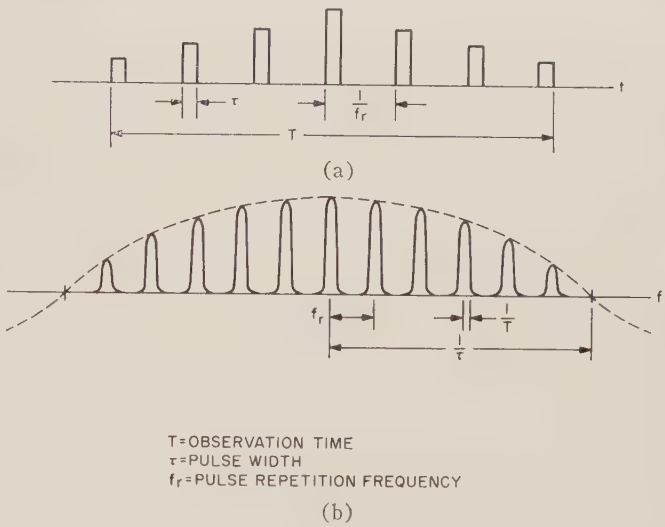


Fig. 7—Characteristics of a pulse train of finite duration.

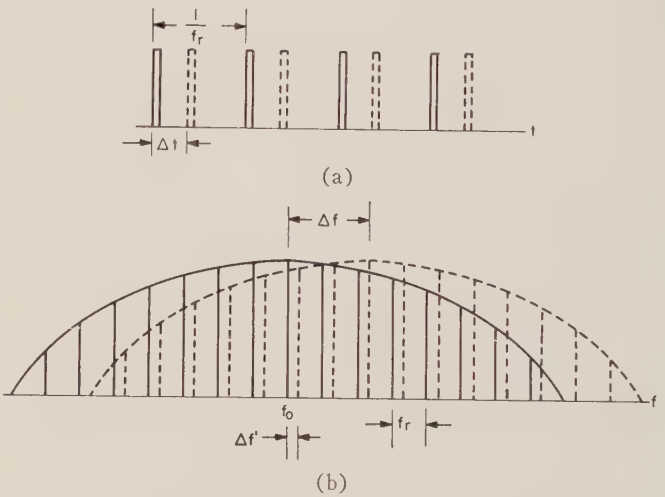


Fig. 8—Transmitted and received signal characteristics.

range could also be measured by frequency-modulating a CW radar.

Fig. 9 shows the time and frequency plots for a CW radar that is frequency-modulated by a single sinusoid. Frequency plots are shown for low and for moderately high modulation indexes. Fig. 9(b) might be an example of a spectrum which affords coarse range measurements with no velocity ambiguities. Fig. 9(c) shows a wider spectrum which might be used to increase the range accuracy while accepting the resultant ambiguities.

A form of radar exists, at least in theory, in which no ambiguities ever occur and where range and velocity determination can be made over any limits with as high a precision as desired. The transmitted signal has a noise-like waveform. The bandwidth and observation time requirements still apply, and the receiving system must be coherent. By remembering transmitted waveform, the reflected energy can be examined by time and frequency comparison with respect to the stored transmission. Only when the time and frequency match is

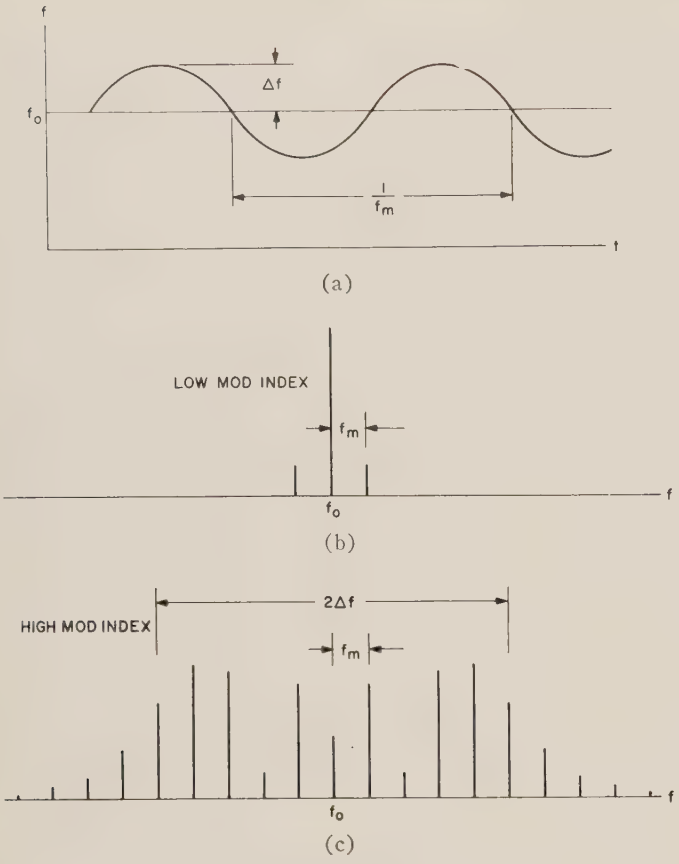


Fig. 9—Frequency modulated CW.

within the accuracy limits designed into the system would a target be detected. Aside from complexity, the main drawback of this system is the lack of resolution capability. Multiple targets would be difficult to locate; in effect, they would tend to jam each other.

Since the ability to discriminate between moving targets and clutter is often the reason for maintaining coherency, three terms which are commonly used as a measure of performance should be defined: clutter rejection, cancellation ratio and subclutter visibility.

Clutter rejection is the measure of a system's ability to attenuate clutter below some specified level. It is usually defined as the maximum clutter amplitude, which after being attenuated does not exceed this level. It is the ratio of the maximum clutter to receiver noise which can be rejected before crossing the specified threshold.

The cancellation ratio is a measure of the clutter rejection available in a cancellation scheme and is defined as the ratio of clutter-to-signal ratios before and after cancellation.

Subclutter visibility is a measure of a system's ability to detect weak targets in the presence of strong clutter. It is specified as the minimum ratio of signal to clutter at which the signal is just discernible. In a pulse system, this ratio is measured for clutter and signal appearing at the same range, while in an unmodulated CW system, it must include the clutter return from all ranges.

Determination of System Parameters

The final determination of radar system parameters is controlled by sometimes conflicting considerations such as performance requirements, economics and form factors. Considering only the performance requirements, three factors—detectability, accuracy and resolution—demand attention in a radar system performance analysis.

Detectability is limited by the ratio E/N_0, where E is the returned signal energy (assuming a finite observation time) and N_0 is the received noise power per cycle per second. Accuracy in range[6] (or time) is limited by the signal bandwidth, and accuracy in speed (or frequency) is limited by the time of observation, both for a constant ratio of E/N_0. Resolution, the capability of distinguishing returns as separate targets, is limited by the spreading and/or ambiguities of the signal in the time and frequency domains.[7] Although it is difficult to formulate a necessary mathematical requirement to guarantee resolution, a sufficient condition is that the spectra or waveforms of the target returns do not overlap. In noncoherent radars, the choice of system parameters is primarily governed by these three factors. In coherent systems, where speed is obtained by direct measurement of Doppler frequencies, ambiguities in frequency and time often present an additional problem.

Consider the case where it is desired to measure both target range and velocity accurately. If a pulse system is used and unambiguous range coverage is required, there will be ambiguities in velocity if the Doppler frequency is greater than PRF/2. If unambiguous velocity coverage is required, ranges greater than $c/4f_d$ will be ambiguous. A frequency-modulated CW radar can avoid this problem. The modulation frequency can be made low enough to allow unambiguous range measurement and the deviation kept low enough to avoid ambiguous velocity. However, a new restriction has now been placed on accuracy, because narrowing the bandwidth of the signal restricts the maximum accuracy of range measurement. A noise-like radar transmission apparently fits all needs, since there are no ambiguities and the accuracy of range and frequency measurement is limited only by the arbitrary bandwidth and the observation time. A restriction in multiple target resolution capability now limits performance.

It appears that there is no approach to the design of a coherent radar which will allow opportunity for unlimited accuracy, resolution and freedom from ambiguities. This has been confirmed by the theoreticians and is a consequence of "The Radar Uncertainty Principle."[7,8] Given a specified waveform, the probability of

[6] The determination of target angular positions is usually similar for both coherent and noncoherent radars; therefore the treatment of this problem has been omitted.

[7] P. M. Woodward, "Probability and Information Theory, with Applications to Radar," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955.

[8] W. M. Siebert, "A radar detection philosophy," IRE Trans. on Information Theory, vol. IT-2, pp. 204–221; September, 1956.

target detection by a matched filter can be determined as a function of the filter time and frequency coordinates with respect to those of the target. The value of this detection probability, when squared and integrated over all time and frequency is a constant, independent of the transmitted waveform. Woodward[7] and Siebert[8] have demonstrated that detectability, accuracy, ambiguity and resolution are interdependent.

To illustrate some of the principles stated earlier, consider an application of a coherent radar system to an extreme case, such as tracking a satellite, when the following detailed requirements apply: maximum range—700 nautical miles, maximum velocity—25,000 ft/sec, range accuracy compatible with 1 Mc bandwidth.

Three types of systems will be considered: a high and a low PRF pulsed Doppler radar and an FM CW radar, all at *C*-band.

The high PRF radar would require a PRF of 290 kc merely to avoid blind velocities and double this to avoid ambiguous velocities. At 290 kc, blind ranges would occur every 1700 feet. For a fixed PRF the target must compete with returns at all ranges and fly through the blind ranges at a maximum rate of 15 per second, causing 15-cps amplitude modulation of the signal. Modulation products would appear as 15-cps harmonics around each frequency component in the return spectrum, thus restricting the accuracy with which the Doppler frequency may be measured. A scheme to avoid the passage through blind ranges would cause considerable complexity in equipment design.

The low PRF system with unambiguous range would require a PRF of about 115 cps, with resulting ambiguous speeds every 5 ft/sec. Any useful clutter rejection would be out of the question, since rejection of clutter modulations from 0 to 5 ft/sec would eliminate all target returns, moving as well as fixed. Even if very low velocities were to be rejected, any acceleration of the target would again cause amplitude modulation of the signal as it passed through the blind velocities. As with blind ranges for the high PRF system, this would spread the PRF lines, and at such a low PRF might even cause the lines to overlap.

Similar phenomena occur with the use of an FM CW system. In order to meet the range requirement, a bandwidth of 1 Mc (approximately equal to $2\Delta f$ or $2mf_m$) will be necessary so that mf_m must be $\geqq 500$ kc. A choice comparable to that between a high and a low PRF pulsed Doppler radar must be made, and ambiguities must be accepted.

For unambiguous range to 700 nautical miles, f_m cannot be greater than 115 cps; therefore, the resulting separation of lines in the frequency spectrum will create speed ambiguities every 5 ft/sec. For unambiguous range rate to 25,000 ft/sec, f_m must be $\geqq 580$ kc (twice the expected Doppler frequency shift). This results in range ambiguities every 850 feet. At values of f_m between these extremes, there are ambiguities in both range and range rate.

A CW system with a large mf_m product is possible, but it loses the basic advantages of simplicity and narrow bandwidth without gaining the advantages of a pulsed Doppler system, particularly the high rejection of targets outside the range gate.

The previous examples demonstrated that simple radar would have difficulty with the simultaneous application of extreme range, speed and accuracy requirements. Depending on the costs, form factors, etc., the choice of a final system lies between coping with the shortcomings outlined or applying the equivalent of multiple radars side by side, which can, in effect, break down the requirements into wieldy groups. Most system requirements are obviously much less severe and the resulting system designs are usually quite reasonable.

There are many applications where the range accuracy requirement is of secondary importance, such as beam rider or semi-active homing missile systems. In this case, the FM CW radar will be a natural consideration, since unambiguous range and velocity may be had at the expense of range accuracy.

In some amplications where pulse radars are the natural choice because of tighter range requirements, it often becomes necessary to use a system with ambiguities. If target velocity is important only as an indication of target motion, then ambiguities in velocity may be disregarded. However, when the Doppler shift of the target return is equal to some multiple of the PRF, blind velocities will occur, and the system design should include a method of avoiding them.

Range ambiguities can be tolerated without any special considerations if a means of determining coarse target range is available. If the receiver system is blind during transmit time, it may be necessary to develop a technique for avoiding the blind ranges at n/PRF intervals.

The obvious method of avoiding blind speeds and ranges is to vary the system PRF. Fig. 10(a) indicates the frequency pass-band characteristics of the coherent pulse radar using a range-gated clutter-filter technique at two different PRF's. Note that the blind frequency bandwidth is equal to twice the clutter rejection frequency, f_c. Switching between these two PRF's will allow coverage of Doppler frequencies up to the frequency where there is an overlap in the blind region for both PRF's. Fig. 10(b) shows the blind range intervals for two PRF's, and similarly it can be seen that the combination of the two PRF's allows unrestricted coverage out to a range where the blind regions overlap. If both blind ranges and blind velocities are to be avoided simultaneously, then more than two PRF's will be required.

PRF variation may be used to solve ambiguities in range and velocity as well as to avoid blind spots. Because the actual target range is equal to the apparent or ambiguous range plus some multiple of the transmitted pulse interval, true range can be determined by observations at two PRF's. Likewise, the true Doppler fre-

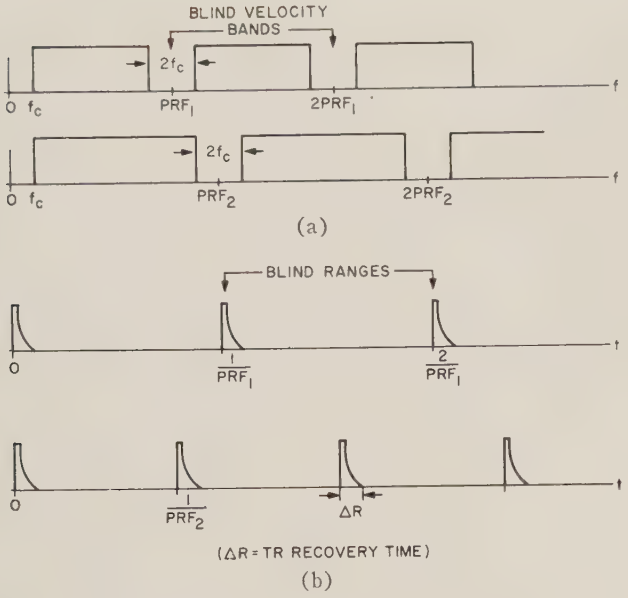


Fig. 10—Dual PRF's for avoiding blind regions.

quency can be determined by measuring the ambiguous Doppler frequencies at different PRF's. Schemes for rapid solution of ambiguities which require modulation of the PRF should be considered with some caution, since modulation of the PRF will cause additional lines to appear in the frequency spectrum and may hinder the resolution of the Doppler frequency.

Another system parameter which affects the ambiguity is the transmitter frequency. If it is desired to maintain unambiguous range and Doppler frequency measurement, the following relationship between PRF, transmitted frequency and maximum unambiguous velocity may be exploited:

$$\frac{\text{PRF}}{f_0} = \frac{4v_{\max}}{c}.$$

Decreasing the transmitted frequency will provide more Doppler space for unambiguous velocity measurements without affecting the range PRF requirement.

Coherent system techniques allow target detection at weak signal levels lower than those detectable with noncoherent techniques. One such technique is the use of narrow-band filters before envelope detection in order to reduce the signal-to-noise loss caused by noise intermodulation.[9] Serious consideration should be given to the relative merits of increasing transmitter power or receiving system complexity in order to achieve the desired system detectability.

Two other important considerations must be included under the subject of system parameter determination for coherent radars: subclutter visibility and clutter rejection. Although clutter is not always a significant problem, it generally requires consideration where CW systems are involved, or where pulse systems for short-range application or having ambiguous ranges are used.

Subclutter visibility, as previously defined, is a measure of the system's ability to detect small targets in the presence of large clutter. In CW systems or ambiguous pulse systems where a continuous CW reference is used, the limitation in subclutter visibility is usually determined by the amount of noise on the transmitted signal. When the target return is accompanied by clutter, the noise in the received signal will include the transmitter noise at an amplitude proportional to the clutter amplitude. Thus there is an obvious limitation in the ability to detect a target of given cross section in the presence of clutter. In a CW system, range performance will be affected by the degree of subclutter visibility and local clutter conditions since all the clutter will compete with the returned target. In an ambiguous-range pulse system, the range will be somewhat affected by subclutter visibility and clutter conditions because it may be necessary to observe long-range targets buried in ambiguous short-range clutter.

Clutter rejection in coherent systems is accomplished either with cancellation techniques for pulse-to-pulse comparison systems, or with filters in CW and pulsed Doppler systems. The worst result of inadequate clutter rejection is usually the production of false target alarms. If clutter rejection is achieved after phase comparison, an additional problem of harmonic distortion arises when moving clutter is encountered. When harmonic distortion in the circuitry between the phase detector and the clutter rejection filters is excessive, multiples of the clutter frequency may appear in the target pass band. This is a situation which might be encountered when the radar is on a moving platform (such as a ship or a tank) and is one reason why clutter rejection at IF may be preferred.

Target Analysis in Coherent Radars

Target analysis is usually considered to be the evaluation of target parameters such as position, speed, course, size, maneuvers, etc. Unfortunately, even perfect resolution of targets in space and frequency is not sufficient to meet the requirements of many tactical situations. Some new techniques are in order. Serious effort has been directed toward detailed spectral analysis of the target return. This analysis may amount to the examination of target modulations from which useful information such as target engine type, size, number of vehicles, detection of maneuvers, decoy identification and countermeasures analysis is available to some extent.

The advent of operational coherent systems has been of particular aid in the advancement of these spectral techniques because of the availability of the modulation information in a usable form. In a noncoherent radar, the target modulations are irretrievable except for those components which are strictly amplitude modulation. The reason for this is evident when the modulations are

[9] R. A. Smith, "The relative advantages of coherent and incoherent detectors: A study of their output noise spectra under various conditions," *Proc. IEE*, vol. 98, pp. 43–54; October, 1951.

considered to be the complex resultant of various signals. If the signals are compared, the relative phase from pulse to pulse is lost because the noncoherent radar, as explained previously, has either no continuity in phase or no means of storing the transmitted pulse phase for comparison with the signal return. Because a coherent system maintains a continuous phase reference, the return signal will exhibit all the target modulation frequencies when they are detected and filtered.

Once the target modulations have been recovered, they can be visually displayed as a spectrum or aurally presented. The aural display is the simplest to obtain since it consists merely of an audio system with the proper band-pass characteristics. The extent to which this can be used will depend somewhat on the system transmitted frequency and target speed if unambiguous Doppler is to be used, because the Doppler frequencies might be beyond the audible range. The acceptance of ambiguous Doppler frequencies could solve this problem.

An aural display also allows the operator to perform as a narrow-band detector since the signal-to-noise ratios detectable by the ear are quite good and in the order of -15 db. The audio target "signature" of a jet aircraft is often characterized by a tone which is not a direct function of the target's speed, but rather a function of the turbine RPM. Propeller-driven aircraft also have their unique audio signatures.

It is conceivable that more sophisticated means of target modulation spectrum analysis could be made with some means of automatic analysis, but it is unlikely that a simple device could be made to duplicate the performance of a trained operator.

The Counter-Countermeasures Aspect

Little can be said in this paper about the counter-countermeasures aspects of the coherent radar because of security considerations. However, the clutter rejection and subclutter visibility properties of coherent systems are obvious advantages against the chaff types of countermeasure. Because the trained operator is still an excellent counter-countermeasure, an additional input to him in the form of audio offers interesting possibilities for the advantageous use of audio target signatures.

Acknowledgment

The authors are indebted to G. J. Plotkin for his helpful criticism and assistance in preparation of the manuscript.

Interferometer Techniques Applied to Radar*

E. GEHRELS† and A. PARSONS‡, member, ire

Summary—**A method is described for measuring the angular rate of a radar target by measuring the rate of change of the phase difference between radar echoes received at two widely separated antennas. It is shown that it is possible by the interferometer technique employed to obtain a direct measure of target angular velocity in much the same manner that Doppler measurements yield radial velocities.**

Three main questions existed concerning the basic feasibility of such an approach:

1) **Will the effect of the fluctuations in time of arrival arising from inhomogeneities in the atmosphere render the data incapable of interpretation?**
2) **Is sufficient phase stability obtainable in the equipment and in the transmission paths between widely separated sites to permit phase measurements of the required accuracies?**
3) **Will serious degradations in the data occur as a result of shifts in the phase center of irregularly shaped tumbling objects?**

* Received by the PGMIL, January 25, 1961.

† M.I.T. Lincoln Lab., Lexington, Mass. Operated with support from the U. S. Army, Navy, and Air Force.

‡ Ewen Knight Corp.; Needham, Mass.; formerly with M.I.T. Lincoln Lab., Lexington, Mass.

Introduction

THE radar interferometer may be defined as a multistatic radar system that measures the difference in the phase of an echo signal as received at two or more stations. For the discussion in this paper it will be assumed that the target is illuminated by one of the stations and observed by all.

The significance of phase difference measurements in a passive radio interferometer can be seen in Fig. 1.

For a distant target, the difference in path length to the two stations of spacing b is

$$l = b \cos \theta. \tag{1}$$

The phase difference between the signals reaching the two stations, assuming a monochromatic illumination wavelength λ, is

$$\phi = \frac{2\pi}{\lambda} b \cos \theta. \tag{2}$$

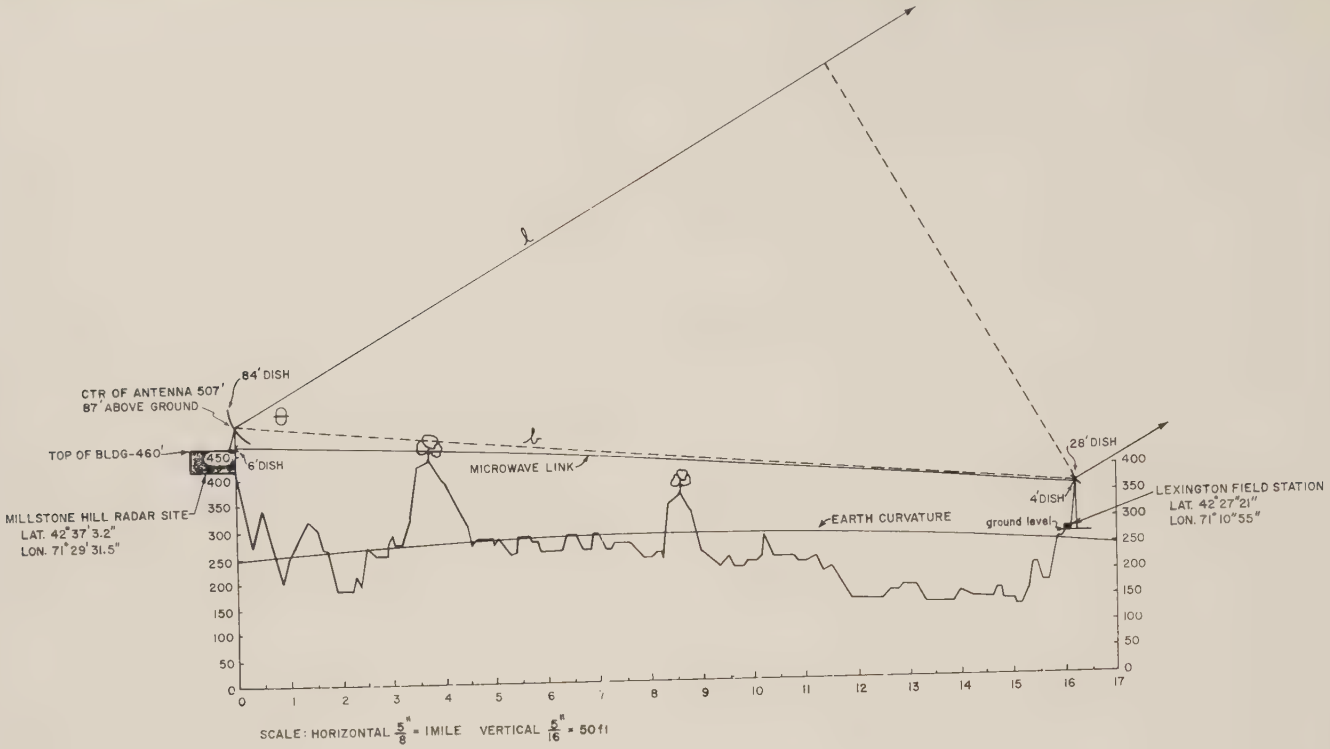


Fig. 1—Profile of Millstone Hill to Lexington field station.

The sensitivity of the interferometer as a function of angle is

$$\frac{d\phi}{d\theta} = -\frac{2\pi}{\lambda} b \sin\theta. \tag{3}$$

Two features immediately become apparent as the spacing of the stations is increased:

1) The value of ϕ becomes an extremely sensitive measure of the polar angle θ.
2) The measure becomes extremely ambiguous, *i.e.*, it is impossible to determine which multiple of 2π radians (360°) of phase angle corresponds to the phase path difference.

The situation is analogous to having a clock missing all the hands but the second hand; one could read the time very accurately, but highly ambiguously. However, to carry the analogy further, if this clock were continuously observed, it would be possible, by keeping a count of the number of revolutions, to achieve an unambiguous measure of the elapsed time between any pair of events. Thus, the angular velocity of a target (*i.e.*, the rate of change of angular position) can be measured unambiguously if continuous track is kept of the phase angle.

The radio interferometer first received wide-spread application in radio astronomy. Ambiguities can be resolved by moving one or more of the antennas[1] and making successive measurements, or by the use of multiple base lines to resolve the ambiguities. The minitrack system[2] is a good example of the multiple base line techniques applied to a vehicle tracking problem.

Another possibility for resolving ambiguities is the use of a broad-band signal instead of a single frequency.[3] This approach permits not only the unambiguous determination of direction of a single source but also the resolution of a number of separate sources.

Instead of using a continuous band of frequencies, it is possible to use a discrete set of frequencies or sidebands. Thus, assuming that a carrier frequency is modulated simultaneously at, for example, frequencies of 1/25 and 1/625 of its own frequency, it is possible to measure the relative phases of the detected modulation to achieve coarser measures of phase. This method is used in the AZUSA[4] system at Cape Canaveral, Fla., for resolving phase ambiguities.[5]

From a proposal of Franklin A. Rogers in 1958, Dr. Roger Manasse and John Sheehan at the M.I.T. Lincoln Laboratory designed an experimental two-station interferometer between the main Laboratory and the Millstone Hill radar 25 km distant. At the 440-

[1] R. N. Bracewell, "Radio interferometry of discrete sources," Proc. IRE, vol. 46, pp. 97–105; January, 1958.

[2] J. T. Mengel, "Tracking the earth satellite, and data transmission, by radio," Proc. IRE, vol. 44, 755–760; pp. June, 1956.

[3] L. R. Dausin, K. E. Niebuhr, and N. J. Nilsson, "The Effects of Wide-Band Signals on Radar Antenna Design," RADC, Griffiss AF Base, N. Y., TN 59–178, ASTIA Doc. No. AD 216-023; July, 1959.

[4] AZUSA, Convair Div., General Dyanmics Corp., San Diego, Calif., Convair Rept. No. AZN-004; March, 1959.

[5] In the AZUSA system this technique is only used for measuring the phase difference between the signal transmitted to the rocket and that returned by the transponder, *i.e.*, the range measurement. From practical considerations, a multiple base line is used for the angular measurements.

Mc frequency of the Millstone transmitter, this corresponds to a b/λ of about 38,000. The so-called "fringe width,"[6] or the angular motion corresponding to a 360° phase shift between antennas, varies from 6 seconds of arc near the perpendicular to the base line to a width of 0.4 degree of arc for the first fringe off the end of the base line.

The integration of interferometer techniques with conventional pulsed radar provides a measure of the three coordinates of position plus three velocity components. As was pointed out earlier, it is possible to make an unambiguous measurement of angular *change* over a period if an unbroken count of the phase rotations is maintained. With such a radar system, then, there is no need for removing the ambiguity in the interferometric angular *position* measurement. However, with a pulse radar, the requirement for continuous phase observations cannot be strictly met. In fact, the phase difference may complete a large number of cycles between radar pulses. Also, the relatively low signal strength available (compared to that available with a satellite beacon or transponder) requires a close approach to matched filtering of the signal; this filtering is required *without introducing undetermined phase shifts*. Incidentally, the need for high-gain antennas discourages on an economic basis the use of additional base lines for ambiguity elimination, as might be done in passive tracking interferometers.

Three main questions are listed concerning the basic feasibility of a radar interferometer using such a long base line:[7]

1) Is sufficient phase stability obtainable in the equipment and in the transmission paths between widely separated sites to permit phase measurements of the required accuracies?
2) Will the effect of the fluctuations in time of arrival arising from inhomogeneities in the atmosphere render the data incapable of interpretation?
3) Will serious degradation in the data occur as a result of shifts in phase center of irregular tumbling objects?

The experimental system described here has answered all of these questions optimistically. Work directed at answering each of the questions will be described in turn.

Experimental System

A simplified diagram of the Millstone Radar Interferometer is shown schematically in Fig. 2.

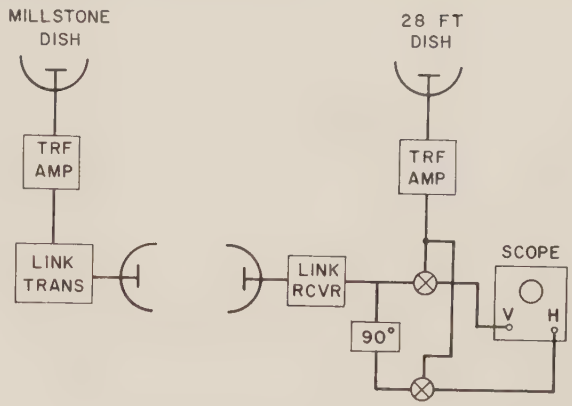


Fig. 2—Simplified block diagram of Millstone-Lexington interferometer.

The Millstone radar typically operates using a simple pulsed signal with a repetition frequency of about 30 per second and a pulse length of two milliseconds. For a low satellite at 200-km range, the maximum angular velocity is 40 milliradians (2°) per second, corresponding to a maximum rate of change of phase difference in the interferometer of 3 revolutions (1080°) per two-millisecond pulse length and 47 revolutions in the interval between the end of one pulse and the start of the next.

If the rate of phase rotation can be measured sufficiently accurately during the two-millisecond pulse, it is easy to extrapolate the number of revolutions undergone between pulses, thus maintaining an accurate count of the total phase shift undergone over a period of time. However, frequently it is not possible to measure the rate of phase rotation, *i.e.*, frequency, so accurately on a single pulse, and doubt may exist whether, say, 47, 48, 49, or 50 cycles of phase rotation have taken place during one interpulse period. This ambiguity corresponds to possible frequency errors in multiples of the repetition frequency, as would be expected from examining the spectrum of a periodically pulsed signal. This, of course, consists of "spikes" spaced at the pulse-repetition frequency with an envelope corresponding to the spectrum of a single pulse. A small amount of noise or equipment instability can cause an adjacent frequency spike to be mistaken for the true center frequency.

Fortunately, a difference of 30 cps corresponds to a difference of $\frac{1}{2}$° in target travel over a ten-second period. With an observation period of this length, the coarse angular rate measurement from the antenna tracking is sufficient to resolve the ambiguity.[8]

[6] This term results from the repeating lobe or fringe-type radiation pattern resulting from an interferometer consisting of two space antennas connected in parallel at the receiver terminals.

[7] At the time of completion in 1959, the base line was, by a substantial margin, the longest known to the authors, both in absolute length and in terms of wavelengths (b/λ). Recently a 40-mile interferometer at 160 Mc has been placed in operation at the Nuffield Radio Astronomy Labs., Jodrell Bank, England and a 1-mile long *S*-band interferometer, using two 90-foot parabola antennas on railway tracks, is under construction at the Royal Research Establishment, Malvern, England.

[8] Another method for resolving the ambiguity involves staggering the pulse interval. For instance, the interpulse interval after the odd-numbered pulses can be made ten per cent longer than normal, while the interval after even-numbered ones is made ten per cent shorter than normal. Without pulse staggering, a 30-cycle frequency difference would appear the same as a zero frequency difference except for a 21° difference in phase shift during the 2-millisecond pulse. The staggered pulse would exhibit a 36° phase advance between pulses 1 and 2, and a similar phase retardation between pulses 2 and 3, a 72° difference between successive phase changes.

The preceding discussion on phase measurements immediately raises questions concerning the effect of noise and optimum methods of processing. If one assumes the noise and other phase-perturbing effects to be completely independent at the two interferometer stations (the latter is not strictly justified), one can show that the optimum filtering consists of independent Doppler filter banks of matched filters at each of the interferometer stations. By measuring the Doppler and phase (with respect to the transmitted frequency) for each station and subtracting, the interferometer frequency and phase differences are obtained.

According to Woodward[9] and Manasse[10] the rms accuracy of a Doppler measurement (angular velocity) of a single rectangular pulse of length T is

$$\sigma_{\dot{\psi}} = \frac{\sqrt{3}}{T\sqrt{R}}, \tag{4}$$

where R is twice the energy contained in the received pulse divided by the noise power per unit bandwidth. (This ratio corresponds to the maximum signal-to-noise ratio available from a properly matched filter and is here assumed to be much greater than one.) For the frequency difference between two stations, with signal-to-noise ratios R_1 and R_2, the rms error is

$$\sigma_{\dot{\psi}} = \frac{\sqrt{3}}{T}\left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{1/2}. \tag{5}$$

The rms phase measurement errors for a single station and for an interferometer are, respectively,

$$\sigma_{\psi} = \frac{1}{\sqrt{2R}}, \tag{6}$$

$$\sigma_{\phi} = \frac{1}{\sqrt{2}}\left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{1/2}. \tag{7}$$

The single-pulse frequency measurement (or, equivalently, the rate of phase change during a single pulse) may be useful in ambiguity resolution; the significance of the phase accuracies will be discussed further.

A simplified block diagram of the processing equipment used for most of the interferometer measurements to date is shown in Fig. 2. The two product (cross-correlation) detectors give outputs proportional to the cosine and sine of the phase difference. The signal filtering is a compromise involving a broad predetection filter and a narrower postdetection low-pass filter matched to the transmitted pulse length. Some such procedure is necessary in a simple system such as this because the use of a conventional bank of nearly optimum predetection filters introduces discontinuous phase shifts as a transition is made from one filter to the next.

The simple system works satisfactorily when the angular rate of the target is small; at high angular rates, when the phase changes by an appreciable portion of a cycle during the two-millisecond pulse length, or when it is desired to measure the phase angle between the transmitted and the received signal at frequencies away from zero Doppler, the system is less than satisfactory. A fair body of experimental data has nonetheless been obtained. In addition, a promising expedient for alleviating the deficiencies of the system discussed here is described in the Appendix.

Equipment Stability

Phase instabilities in the system arise in the radar receivers, in the transmission links and from failure to maintain absolute local oscillator synchronism between the sites. Measurements have shown the receivers to be phase-stable to within $\pm 3°$ over a 40-db dynamic range. The phase variation introduced by the microwave link in transmitting the low intermediate frequency of 200 kc is negligible, of course.

The local oscillator phase reference is obtained from heterodyning a separate link frequency of 407.5 Mc with the 410-Mc local oscillators at each location, as shown in Fig. 3. The local oscillator frequencies are synthesized at each location from an ultra-stable oscillator. The stable source at the Lexington end is manually verniered to maintain a constant phase-difference pattern on the oscilloscope.

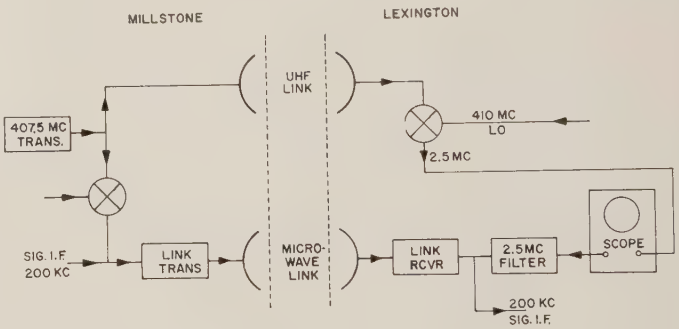


Fig. 3.

The profile of the transmission path between Millstone and Lexington is shown in Fig. 1, with the vertical scale exaggerated fifty-fold. The Westford Hill rise, with its swaying trees, has been an occasional source of phase instability in the microwave link. However, it is believed that the fact that it provides shielding from expanses of level ground has resulted in reduced multipath problems at 407.5 Mc where antenna directivity is much less.

Change in the length of the propagation path and the stability of the 407.5-Mc transmission was measured by a variant of the method used by Herbstreit and Thomp-

[9] P. M. Woodward, "Probability and Information Theory with Applications to Radar," Pergamon Press, Inc., London, England; McGraw-Hill Book Co., Inc., New York, N. Y., chs. 6, 7; 1953.

[10] R. Manasse, "Summary of Maximum Theoretical Accuracy of Radar Measurements" The MITRE Corp., Bedford, Mass., Tech. Ser. Rept. No. 2; February, 1960.

son[11] to measure small changes in propagation time. The 84-foot Millstone dish and the 28-foot Lexington dish were pointed at one another. One watt at precisely 440 Mc was transmitted from Millstone and a similar power at a frequency differing by about 1 kc was transmitted from Lexington.

The "receiver" at each end for this experiment consisted of a crystal diode detector using the outgoing transmitted frequency as its own local oscillator. The phase difference between the 1-kc detector outputs at each end (which were brought together for comparison via a land line) represents the combined phase shifts for a round-trip path. Assuming that the propagation path is reciprocal for the two frequencies, it is possible to observe the variations in the propagation delay. The phase of the 407.5-Mc signal, which is derived from the same standard oscillator, was then compared against the phase at 440 Mc.

Measurements taken with this arrangement and previously reported by Gehrels[12] indicated an over-all phase instability for the 407.5-Mc system of 4° peak-to-peak over a ten-minute period. This corresponds to a relative timing accuracy of 2.5×10^{-5} microsecond.

Obviously, this technique of two-way transmission could be used for compensating instead of merely measuring these errors.

Atmospheric Effects

The troposphere is known to introduce short-term fluctuations in the angle of arrival of line-of-sight radio waves on the order of a tenth of a milliradian.[13] The ionosphere also introduces fluctuations of a magnitude highly dependent upon frequency. From these facts it has been concluded that to build an interferometer longer than one kilometer in the UHF region[14] or to build an antenna with greater than 80-db gain (or even less for antennas pointed at low angles)[15] would be pointless because of the absence of a coherent wavefront over the full aperture. Fortunately, the linear extrapolation of increase in phase shift with increasing base length implied by a constant angular deviation does not hold for very long base lines.[12]

Consider two essentially parallel paths through the atmosphere spaced by distance D. Because of atmospheric turbulence, these paths at a time t offer phase shifts varying from their normal values by $\psi_1(t)$ and $\psi_2(t)$. One may define a cross-correlation function (not normalized) as a function of spacing D:

$$R\psi(D) = \langle \Delta\psi_1(t)\Delta\psi_2(t)\rangle_{\text{av}}. \tag{8}$$

The mean-square variation for one path alone is

$$\langle \Delta\psi(t)^2\rangle_{\text{av}} = R\psi(0). \tag{9}$$

The mean-square phase differences, $\psi_1-\psi_2$, between the two paths are

$$\sigma_\phi{}^2 = \langle(\Delta\psi_1 - \Delta\psi_2)^2\rangle_{\text{av}}, \tag{10}$$

$$= 2\langle(\Delta\psi)^2\rangle_{\text{av}} - 2\langle\Delta\psi_1\Delta\psi_2\rangle_{\text{av}}, \tag{11}$$

$$= 2[R(0) - R(D)]. \tag{12}$$

The result is that at small spacings the mean-square phase difference does increase rapidly with spacing, but at large spacings reaches a limiting value $2R(0)$. Values of these correlation coefficients have been calculated by Muchmore and Wheelon[16] by use of mathematical models designed to be in reasonable agreement with the experimental results of Herbstreit and Thompson.[11] The results indicate at 440 Mc an expected single path phase variation of $3\cdot10^{-2}$ radian and essentially zero correlation with spacings exceeding several hundred feet, the typical size of the turbulent blobs.

Larger-scale effects that in some cases may cause greater total phase changes can result from diurnal and meteorological frontal and layer phenomena.[17] However, these vary slowly enough in time and in position to have very little effect on short-duration velocity measurements. One possible exception is a large isolated cloud of the cumulus variety. A grazing beam entering such a cloud could undergo a 30° phase shift over a few seconds.

The effect of the ionosphere in producing phase fluctuations is not well established. However, the derivation following sets an upper limit to the value of deviations so caused in the angular rates measured with the subject radar interferometer. From short baseline interferometric observations of radio star angular scintillations, Booker,[18] one can infer the maximum differential propagation delay over the short baseline. Keeping in mind that the radar interferometer is measuring angular rates only, rather than angle, and assuming its base line is long compared to the scale size of ionospheric irregularities, one can infer a worst-case error in the measured angular rates, as follows.

[11] J. W. Herbstreit and M. L. Thompson, "Measurements of the phase of radio waves received over transmission paths with electrical strengths varying as a result of atmospheric turbulence," Proc. IRE, vol. 43, pp. 1391–1401; October, 1955.

[12] E. Gehrels, "A Measurement of High Altitude Atmospheric Refractive Effects by Means of a Long Base Interferometer," presented at the URSI-IRE Spring Meeting, Washington, D. C.; May 4, 1960.

[13] W. L. Anderson, M. J. Beyers and B. M. Fannin, "Comparison of computed with observed atmospheric refraction," IRE Trans. on Antennas and Propagation, vol. AP-7, pp. 258–260; July, 1959.

[14] R. S. Lawrence, "An investigation of the perturbations imposed upon radio waves penetrating the ionosphere," Proc. IRE, vol. 46, pp. 315–320; January, 1958.

[15] K. Bullington, "Characteristics of beyond-the-horizon radio transmission," Proc. IRE, vol. 43, pp. 1175–1180; October, 1955.

[16] R. B. Muchmore and A. D. Wheelon, "Line-of-sight propagation phenomena—Section I. Ray treatment," Proc. IRE, vol. 43, pp. 1437–1449; October, 1955.

[17] J. R. Bauer, "The Suggested Role of Stratified Elevated Layers in Transhorizon Short-Wave Radio Propagation," MIT Lincoln Labs., Lexington, Mass., Tech. Rept. No. 124, September 24, 1956.

[18] H. G. Booker, "The use of radio stars to study irregular refraction of radio waves in the ionosphere," Proc. IRE, vol. 46, pp. 298–314; January, 1958.

The angular variation measured by an interferometer of projected length D is

$$\sigma_\theta^2 = \left(\frac{\lambda}{2\pi}\right)^2 \frac{\sigma_\varphi^2}{D^2} . \tag{13}$$

Clearly, the apparent angular deviation decreases with increasing D, as can be seen from

$$\sigma_\theta^2 = 2\left(\frac{\lambda}{2\pi}\right)^2 \frac{R(0) - R(D)}{D^2} . \tag{14}$$

For a short base, the limit may be taken

$$\sigma_\theta^2 = -\left(\frac{\lambda}{2\pi}\right)^2 \lim_{D\to 0} \frac{R(-D) - 2R(0) + R(D)}{D^2} \tag{15}$$

$$= -\left(\frac{\lambda}{2\pi}\right)^2 R''(0), \tag{16}$$

where use is made of the limiting expression for the second derivative and the fact $R(-D) = R(D)$.

The rapidity of the ionospherically-related phase fluctuations for a rapidly tracking beam is determined principally by the rate at which the beam passes from one ionospheric irregularity to the next. The expression for the mean-square variation of the derivative of the phase can be derived in a similar manner:

$$\sigma_{\dot\psi}^2 = -v^2 R''(0), \tag{17}$$

where v is the linear velocity of scan at the layer.

For an angular scanning rate $\dot\theta$ and a layer distance r,

$$v = \dot\theta r, \tag{18}$$

and the rate of the phase fluctuation is

$$\sigma_{\dot\psi}^2 = -r^2\dot\theta^2 R''(0). \tag{19}$$

For a two-station interferometer of projected spacing D, where D is large compared to the scale of the irregularities, *i.e.*, $R(D)$ and $R''(D)$ are zero,

$$\sigma_{\dot\theta}^2 = -2\frac{r^2\dot\theta^2}{D^2} R''(0); \tag{20}$$

$$\left(\frac{\sigma_{\dot\theta}}{\dot\theta}\right)^2 = -2\frac{r^2}{D^2} R''(0). \tag{21}$$

The similarity of (16) and (21) permits deducing angular rate errors from fluctuations in angular position measured with short base interferometers. Booker[18] reports a number of observations of angular fluctuations made of "radio stars" with short base interferometers. Observed scintillations have been typically 0.001 radian at 40 Mc. Scintillations ten times as great have been observed on exceptionally disturbed days (magnetic storms, etc.). Substituting the above equations and correcting for the frequency dependence of the ionosphere gives an angular-velocity error of about 0.15 per cent at high elevation angles. The ionospheric effects are in inverse proportion to the square of the frequency and in direct relation to the cosecant of the elevation angle.

Target Effects

Target scintillation is the effect least amenable to analysis. Although it is intuitively possible to visualize shifts of phase center as an echoing target tumbles, it is not immediately clear whether this will lead to discontinuous changes in relative phase between the two interferometer stations and thus render the data difficult of interpretation. Since analysis and early tests appeared to indicate the feasibility of such a system so far as other questions are concerned, an experimental approach to the target effect question appeared most attractive.

Results of pertinent tests appear in the following sections.

Experiments and Results

An extensive program of satellite observation has been carried out in order to verify the operation of the interferometer. The emphasis has been on the demonstration under different circumstances of the capability to obtain data which are consistent enough to be usable from such a device. The system sophistication necessary to permit ready derivation of orbit parameters from observations has not yet been provided. However, experimental confirmation of feasibility of the approach opens the way for realistic system thinking to this end.

Fig. 4 exemplifies a typical computer plot of phase difference residuals. A best-fit parabola has been subtracted from the raw data in order to make the vertical scale of the plot manageable. Study of the figure shows a value for rms phase error of less than 15 electrical degrees. The observed phase fluctuation results from the combined effects discussed, but excludes possible slow tropospheric and ionospheric factors that could result in bias errors extending below the period during which a parabola matches the geometry of the experimental situation. Fifteen electrical degrees corresponds to a positional movement of 1 meter at 1000 km!

Fig. 5 shows a target in transition from a comparatively steady to a highly scintillating signal. The scintillations were sufficiently severe to cause difficulties for the Millstone automatic tracking system. The phase, however, can still be followed consistently, albeit with added "noise."

The observed phase jitter is greater than that explained by atmospheric effects alone. A strong component having a periodicity one fourth the PRF is clearly visible and is an effect of the conical scan feed system of the Millstone tracker. The remaining instabilities must be ascribed to the other causes discussed earlier.

The above results are representative of those from many targets observed with the system, including Sputnik V (rocket case), the Echo balloon, payloads and

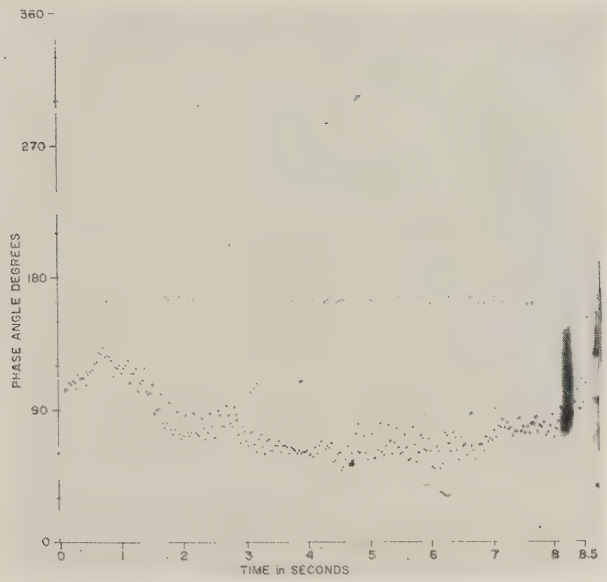


Fig. 4.

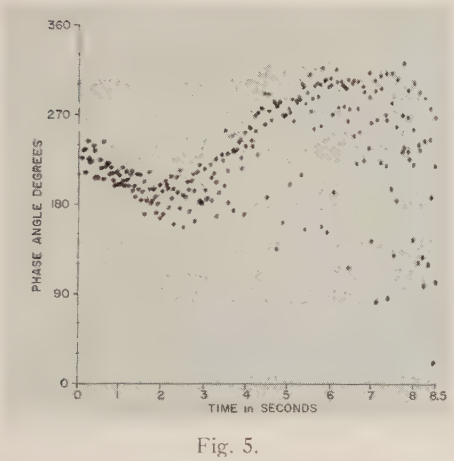


Fig. 5.

rocket cases of the TIROS and TRANSIT series, and several Discoverer (AGENA) vehicles. Target scintillations, though causing a noticeable degradation in phase stability, do not appear to render the data unusable for orbital determination—at least for the objects observed. Thus, the feasibility of a radar interferometric approach to this sort of measurement seems to be established.

Some System Considerations

The radar interferometer was conceived as a means for supplementing the radial (Doppler) velocity and position measurements from a monostatic radar with measurements of the transverse velocity components in order to permit the establishment of a complete velocity vector at a point in space. It now appears that the interferometer velocity measurements are more accurate than the conventional Doppler measurements as now made.

The phase-measurement technique used in the interferometer appears directly applicable to the extremely accurate measurement of radial velocity, or what might be called "fine Doppler." It is possible that target scintillation effects could defeat the fine Doppler measurements where they do not greatly affect the measurement of the phase *difference* between two stations. Preliminary results, however, are encouraging in this respect, and, in the system calculations to follow, it is assumed that the radial velocity can be measured at least as accurately as the transverse velocities.

If the values of phase error are statistically independent for each of the pulse returns (as is the case for contributions from receiver noise), the phase rate error is given approximately by

$$\sigma_{\dot{\phi}} = \frac{2\sqrt{3}\sigma_{\phi}}{\sqrt{Nt}}, \tag{22}$$

where

σ_{ϕ} is the rms deviation of the phase,
N is the number of pulses,
t is the total time of observations.

If the observational errors are not independent (as is the case in periodic fluctuations due to target scintillation, atmospheric inhomogeneities, tracking errors, etc.), the maximum error may be the considerably larger value

$$\sigma_{\dot{\phi}} = \frac{2\sigma_{\phi}}{t}. \tag{23}$$

Using the more pessimistic formula, $\sigma_{\dot{\phi}}$ may be as large as 0.25 radian per second or 0.08 cycle per second for a two-second observation.

Based on the above results regarding electrical phase stability, the accuracy of the angular-velocity measurement of a two-base-line interferometer can be derived as follows. Consider an interferometer with two equal-length base lines, one east-west and the other north-south. Consider also a target elevation α and azimuth β at a slant range large compared to the base-line length. Let ϕ_1 and ϕ_2 be the phase difference angles from the two base lines, respectively. Then,

$$\dot{\alpha} = -\frac{2\pi b}{\lambda \sin \alpha}(\phi_1 \cos \beta + \phi_2 \sin \alpha), \tag{24}$$

$$\dot{\beta} = \frac{2\pi b}{\lambda \cos \alpha}(\phi_1 \sin \beta - \phi_2 \cos \alpha). \tag{25}$$

Setting $\sigma_{\phi 1} = \sigma_{\phi 2} = \sigma_{\phi}$, the rms deviations for α and β are

$$\sigma_{\dot{\alpha}} = \frac{2\pi b}{\lambda \sin \alpha}\sigma_{\dot{\phi}}, \tag{26}$$

$$\sigma_{\dot{\beta}} = \frac{2\pi b}{\lambda \cos \alpha} \sigma_{\dot{\phi}}. \quad (27)$$

In practice, one would make one station common to the two base lines. In this case, the errors in the two interferometers are no longer independent and it can be shown that optimum arrangement is to place the stations in an equilateral triangle. The errors are still given by (26) and (27).

The important result is that the accuracy of angular velocity measurements depends only upon the elevation angle of the target, not upon whether its direction is favorably situated with respect to one or the other of the base lines. At 5° elevation, the accuracy of the azimuthal rate would be 1.0 microradian per second and the measurement of elevation rate would be 3 microradians per second.

It has been computed that, using a pair of interferometers having 25-km base lines, a two-second observation of a satellite of low orbital eccentricity passing to within 1000 km permits measurements of the position of any point on the orbit to within 10 km accuracy and the orbital period to within 10 seconds. The values of the accuracies of the positional measurements used were those of the Millstone radar, 0.15° accuracy in elevation and azimuth, 5 km in range. In fact, it appears that the radar positional measurements constitute the principal accuracy limit to the orbital measurements.

Appendix

The predetection filtering used on the most recent experiments provides the theoretical limit in signal-to-noise ratios. This is accomplished by means of a Simoramic Analyzer,[19] a clever device that uses recirculating delay lines to provide filtering that is continuous in frequency (rather than at discrete frequencies as in a Doppler filter bank). The phase recovery methods are similar in principle to those described, but the processing bandwidths must be thousands of times as great. Tests have just been completed in processing the receiver data in real time in the CG 24 computer at Millstone. The data so far indicate complete success in this method of processing.

Acknowledgment

The authors appreciate the help received from G. M. Hyde and E. N. Dupont in constructing the equipment used in this experiment, and the assistance and encouragement of P. B. Sebring in the preparation of this article.

Appreciation is expressed to the Columbia University Electronics Research Laboratory and to the Rome Air Development Center for making available, on loan, the Simoramic Analyzer.

[19] Federal Scientific Co., New York, N. Y., Rept. T-1/100; October 15, 1958.

New Techniques in Three-Dimensional Radar*

MURRAY SIMPSON†, MEMBER, IRE

Summary—**The definition and basic equations defining the performance of three-dimensional radar systems are given. In particular the parameters defining data rate for different classes of three-dimensional radar are analyzed. Three-dimensional radars are divided into three classes as follows: single beam-rapid scan systems; multiple beam-scanning systems; and multiple beam-nonscanning systems. Each of these types is defined and compared in terms of important characteristics. It is shown that new developments in electronic scanning and multiple beam antennas have made feasible many of these new three-dimensional radar systems. A number of the more important antennas and their method of utilization in the three-dimensional radar equipment are described. Examples are given of two types of three-dimensional radar that are typical of two classes of this equipment. Finally, some of the more important applications for modern three-dimensional radar equipment are noted.**

* Received by the PGMIL, January 24, 1961.
† Maxson Electronics Corp., New York, N. Y.

I. Introduction

In recent years there has been a substantial amount of interest in the development of three-dimensional radar systems. The need for these systems has been made more apparent by the rapid increase in speed and other performance characteristics of airborne targets, such as high-speed aircraft, missiles, and satellites, as well as the continuing need in military areas for obtaining better search and track data on conventional artillery and mortar projectiles. Recent advances in the art of rapid scanning and multiple beam antennas, as well as better techniques in radar-data handling and processing have made the new three-dimensional radar systems feasible.

A three-dimensional radar is defined as a system which can provide continuous information of three in-

dependent spatial coordinates, such as azimuth, elevation and range, and in this manner identify the position of the target in space. This is contrasted to a conventional two-dimensional radar which provides continuous information on only two of these coordinates. The important term in this definition of the three-dimensional radar is the word "continuous." This implies that the spatial information on all targets within the volume covered by the radar be provided in a time interval which falls within the desired or required system-data rate. Thus, a conventional pencil beam-tracking radar would not fall within the definition of a three-dimensional radar by virtue of the fact that the time interval required to cover a reasonable space volume, such as a hemisphere or a substantial portion thereof, would be far in excess of the maximum time permitted to obtain sufficient information on high-speed target coordinates in order to make use of this data for tracking purposes. This time interval, known as the data rate of a system, is an extremely important parameter which forms a basic part of the specification for any three-dimensional radar and frequently governs the specific design of this equipment.

The other important parameters are those which are well known to the radar designer, such as radar range, transmitter power, beamwidth, accuracy and resolution, integration time, requirements for moving target indication and/or target velocity data.

II. Radar Parameters

In order to determine better the basis for design of a three-dimensional radar, it is desirable to review some of the basic equations governing radar performance. The basic radar equation[1] is

$$P_r = \frac{P_t G^2 \sigma \lambda^2}{(4\pi)^3 R^4}, \tag{1}$$

where

P_r = power of radar receiver,
P_t = radar transmitter power,
G = radar antenna gain,
σ = target cross section,
λ = wavelength,
R = target range.

For a specific radar system having a pulse width τ and a total loss L in the RF transmission line, as well as a receiver noise figure $\overline{NF}$, the receiver signal-to-noise ratio is given by

$$\frac{S}{N} = \frac{PG^2\lambda^2\sigma}{(4\pi)^3 R^4 L \overline{NF}(KT)\left(\frac{1.2}{\tau}\right)}. \tag{2}$$

The probability of intercept, more commonly known as the blip scan ratio, is defined as the probability of seeing a target on any one radar look. This derives from the fact that noise is a statistical parameter. Tied in very closely with the probability of intercept is the false-alarm rate. This is defined as the time interval before a noise pulse will exceed the signal plus noise for a particular threshold setting of the radar receiver. Fig. 1 shows the relationship between the single-pulse probability of detection and the signal-to-noise ratio for nonfluctuating targets and for fluctuating targets for various false-alarm rates.[2]

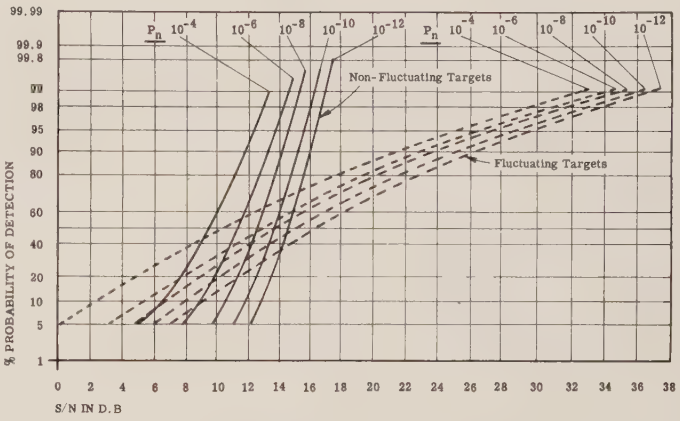


Fig. 1—Single-pulse probability of intercept vs SNR for various false-alarm rates.

The data rate of the radar system is given by

$$D = \frac{\theta \times \mathrm{PR}\overline{F}}{VN}, \tag{3}$$

where

θ = antenna solid angle,
PRF = radar pulse-repetition frequency,
V = total solid-angle scan volume,
N = number of pulses per look.

The parameter θ can be composed of a single antenna beam or the composite of a number of multiple antenna beams. The necessary antenna beamwidth θ is determined by the required angular accuracy and resolution for the radar system, as well as by the range performance as given by (1) and (2). In (3), it can be seen that for a given angular accuracy, the data rate can be increased either by reducing the scan volume or, preferably, by increasing the number of simultaneous beams in the radar system. The latter approach is most commonly used in three-dimensional radar.

III. Types of Three-Dimensional Radar

We will now consider the various general classes of three-dimensional radar systems which can be divided into three categories as follows:

[1] L. N. Ridenour, "Radar Systems," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1.

[2] W. Hall, "Prediction of pulse radar performance," Proc. IRE, vol. 44, pp. 224–231; February, 1956.

Class 1: Single beam-rapid scan systems.
Class 2: Multiple beam-scanning systems.
Class 3: Multiple beam-nonscanning systems.

A Class 1 system is similar to a conventional radar with the exception that the antenna beam can be scanned very rapidly over the required volume. This had not been possible until the advent of electronically scanned antennas, which made it possible to move the beam rapidly in either one plane or two planes without any mechanical motion of the basic antenna structure. Section IV of this paper will describe in more detail different types of antenna systems which satisfy this requirement.

The major limitation in the Class 1, three-dimensional radar derives from limitation on data rate, as given in (3). It can be seen that this system is most readily applied to shorter-range applications where the PRF can be made high and also where the number of pulses required to be integrated is small. Otherwise, the data rate tends to suffer and the system can no longer be considered a true three-dimensional radar. As indicated above, the electronic scan can exist in one plane, such as in the height-finder application in which the beam is scanned electronically in elevation, while the azimuth scan is obtained by normal mechanical rotation of the antenna. Alternatively, the system can be designed so that the beam is scanned electronically in both azimuth and elevation by means of one of several types of two-dimensional electronic-scan antennas described in Section IV.

A Class 2, three-dimensional radar makes use of a multiplicity of simultaneous beams in one plane, the entire group of which is scanned in the orthogonal plane. A block diagram of a system of this type is shown in Fig. 2. The necessary data rate in a Class 2, three-dimensional radar is obtained by setting the value of θ in (3) equal to the total solid angle of n beams in the antenna, as shown in Fig. 2. Thus, all other conditions being equal, the data rate of the system is equal to n times that of a Class 1 system while, at the same time, the accuracy and resolution is given by θ/n, or the beamwidth of an individual beam in the system. It can be seen from Fig. 2 that separate receiver channels are provided for each of the beams of the antenna. Thus, it is not only necessary to provide parallel beams in the antenna, but also a set of n channels in the receiving system. The number of transmitters employed in this system may range anywhere from one to n, depending on the required performance and degree of allowable complexity.

It is interesting to note in Fig. 2 that a requirement exists for a beam selector. This is due to the fact that it is necessary to determine in which of the n beams and receivers the target or targets are being received. This differs from a Class 1 radar in which this information is obtained primarily by the scanning information which continuously provides an indication of the beam position in space.

A typical technique that is employed for beam selection for systems of this type is shown in Fig. 3. A set of three adjacent beams are indicated, namely, N, $N-1$, and $N+1$. These beams are generally designed to overlap at the 3-db point. The circuit is designed so that an attenuated output (approximately 10 db) for each of the three beams is given by AN, $A(N-1)$, and $A(N+1)$. In the system described, it is possible to provide information of target position corresponding to the center of the beam, known as the peak region, and midway between two adjacent beams, known as the junction region. In other words, twice as many possible target positions exist as there are numbers of beams. Further circuit refinement can provide additional possible target positions. Thus, a target would be registered at the peak position at beam N if the output of the AN channel exceeded the output of the $N+1$ and $N-1$ channels. If the outputs are received in N and $N+1$ channels, but do not satisfy the requirements for registration in the peak of either of these two beams, they are automatically assigned to the junction region. In this manner, it

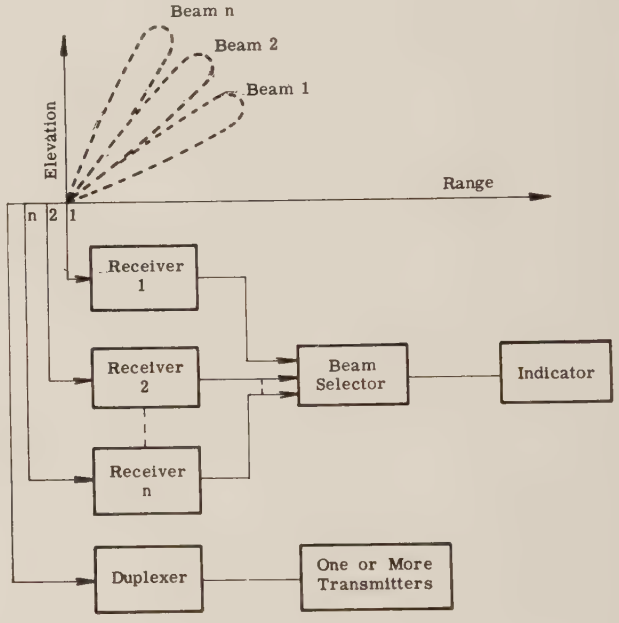


Fig. 2—Typical multiple-beam radar system.

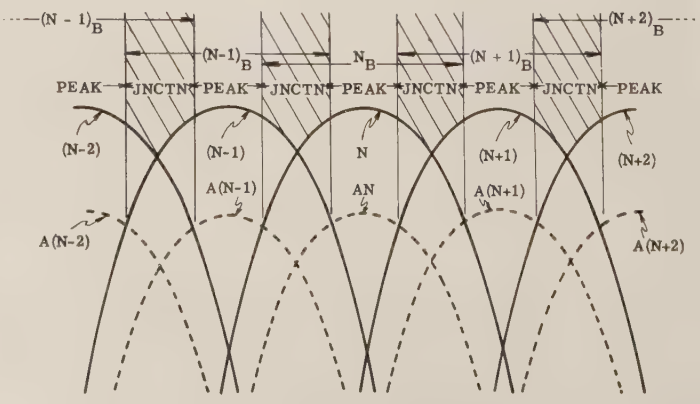


Fig. 3—Target location in multiple-beam radar.

is possible to identify instantaneously the position of targets within the multiple-beam complex on a pulse-by-pulse basis, without the necessity for beam scanning.

It should be noted that the system requires gain stability in the parallel receiver channels in order to prevent errors in beam selection. The receivers used in a system of this type are normally logarithmic in order to provide the necessary dynamic range. In addition, they must be closely matched to a tolerance of approximately 1 db in order to minimize positional errors.

Although the normal Class 2 radar system provides antenna beams which are spaced approximately 1 beamwidth apart, in order to provide solid coverage over the angle θ, there have been systems designed which combine Class 1 and Class 2 operations by means of separating the beams by a number of individual beamwidths and scanning electronically in the plane of the beams, as well as in the orthogonal plane. This type of design is desirable where the required individual beamwidth is extremely small and it is simply not feasible to provide a total number of separate beams spaced one beamwidth apart to cover the required angle θ.

A Class 3, three-dimensional radar is composed of simultaneous beams which do not require any scanning over the required coverage volume. Fig. 2 can be used to illustrate a Class 3 radar, with the understanding that the n beams exist simultaneously over the required coverage volume, generally in two dimensions. This type of antenna coverage can be provided by several types of antennas described in Section IV. Although Class 3 radar provides maximum data rate, since for this system $\theta = V$ in (3), it also involves the greatest system complexity, since it obviously requires a very large number of beams and receiver channels and, possibly, transmitter channels for a reasonable accuracy requirement. Class 3 radars have found particular application for requirements where the total radar coverage volume V is limited so that the resultant number of beams is not too unreasonable. This system is also often used in combination with Class 2 radar. In this case, a two-dimensional cluster of beams is scanned through the total coverage volume V. The value of θ in (3) is then determined by the total solid-angle coverage of the two-dimensional beam cluster.

A comparison of the important characteristics for the three classes of three-dimensional radar is given in Table I.

IV. Antenna Systems for Three-Dimensional Radar

It has been indicated that the feasibility of three-dimensional radar has been determined to a large extent by the development of electronically scanned and multiple-beam antennas. This has been made possible by the ability to move a radar beam orders of magnitude more rapidly than could be accomplished by the mechanically rotating mounts that are used with conventional radar antennas. The technique that is employed in electronic scanning is actually based on the early principles used in directive antenna arrays.

As seen in Fig. 4(a), a pair of radiating sources fed from a common transmitter can be made directive at the angle θ, providing the phase of one element with respect to the other is set at $\phi = 2\pi d/\lambda \sin\theta$, where d is the spacing between elements and λ is the wavelength. This can readily be extended to n radiating elements of a linear array. For the series-fed array in Fig. 4(b) the phase at each element is given by

$$\phi_m = \frac{2\pi}{\lambda} \sum_{x=1}^{m} d_x \sin\theta - \sum_{x=1}^{m-1} \phi_x. \tag{4}$$

TABLE I

	Data Rate	Volume of Coverage	Antenna Complexity	Receiver Complexity	Transmitter Power Capability
Class 1	low	medium	medium	low	low to high*
Class 2	medium	high	medium	medium	medium
Class 3	high	low	high	high	high

* Depending on specific configuration; *i.e.*, whether a single transmitter is used for entire system, or individual amplifiers for each element of antenna.

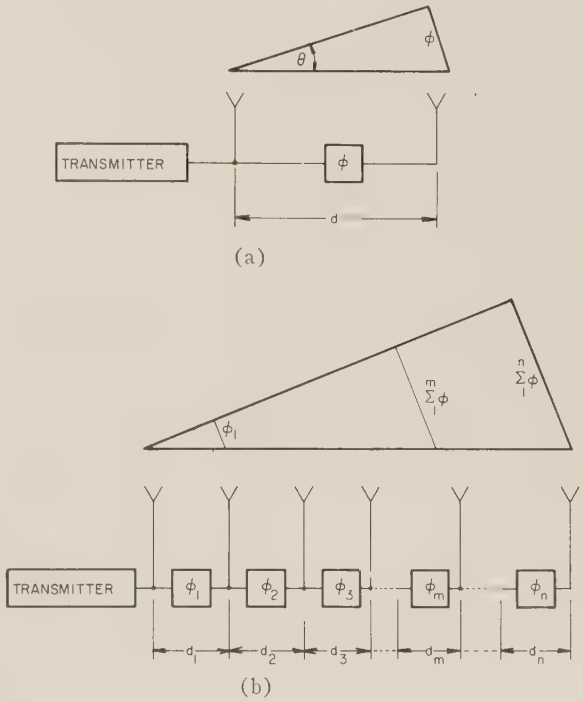


Fig. 4—(a) Two-element series-fed array. (b) n-element series-fed array.

For equal spacing between elements, this reduces to

$$\phi_1 = \phi_2 = \phi_m = \phi_m = \frac{2\pi d}{\lambda} \sin \theta. \quad (5)$$

For relatively long arrays (in wavelengths), the beamwidth is approximately equal to λ/d radians, depending on the amplitude distribution. In order to avoid grating lobes in spurious directions, it is necessary to maintain spacing between elements approximately $\lambda/2$ or less.

It can be seen from the above discussion that in order to obtain electronic scanning of the antenna beam, it is simply necessary to vary electronically the phase at each radiating element. This can be done in various ways. Perhaps the most straightforward technique for varying the phase is simply to change the frequency of the transmitter, since the phase at any point in an RF transmission line is proportional to frequency.[3] This will produce a phase variation at each element given by $\Delta f(L/\lambda)$, where Δf is the proportional frequency change and L is the length of transmission line between radiating elements. An array of this type is known as a frequency-scanned linear array, in which the rate of scan is determined by the rate of frequency change, which can be made as high as desired by the use of electronically-tunable transmitting tubes.

The system shown in Fig. 5 makes use of this principle in combination with a set of orthogonal phase shifters $\phi_1 - \phi_n$, to produce a two-dimensional phase-frequency, electronically-scanned radar system. The linear arrays that are used in this system are known as serpentine arrays, due to the fact that the transmission lines wind around between radiating elements in order to produce a high L/λ ratio, thus providing the desired phase sensitivity for relatively small-frequency change. The set of n arrays make up the complete antenna. The phase shifters ϕ_1 to ϕ_n are connected at the input of each array, and provide beam scanning in the orthogonal plane to that produced by the frequency change. The phase-shifting elements that are employed in this system are ferrite devices whose phase is a function of the magnetic field applied to the ferrites.[4] Thus, in the system shown in Fig. 5, beam scanning in the horizontal plane is provided by the frequency-scanning transmitter, while independent beam scanning is provided in the vertical plane by the change in magnetic field determined by the antenna-beam computer. The position of the beam must, therefore, be determined by noting the specific transmitter frequency and the ferrite-field current. Limitations of this type of system are caused by the necessity to provide very accurate calibration of the ferrite phase as a function of current (a quantity which is generally highly sensitive to temperature), as well as the necessity to make use of a relatively broad-frequency band in the transmitter, thus occupying a substantially larger spectrum than would be required for the normal radar-pulse width.

An alternative system is shown in Fig. 6. In this case,

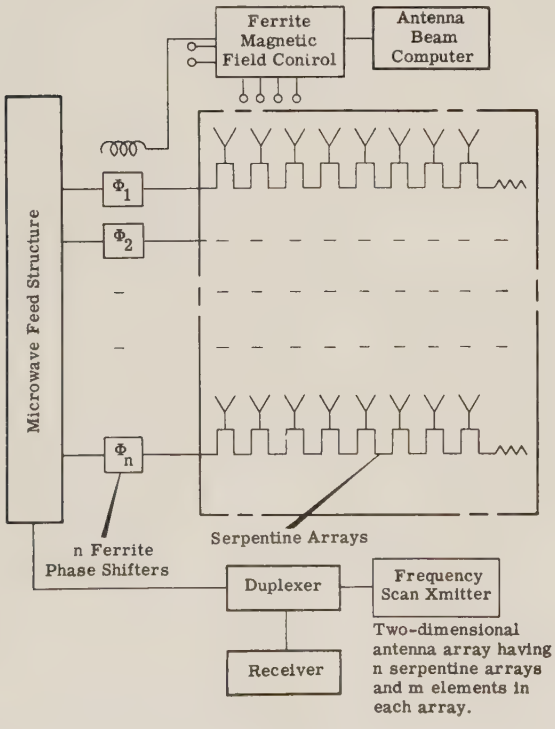


Fig. 5—Phase-frequency electronically-scanned radar system.

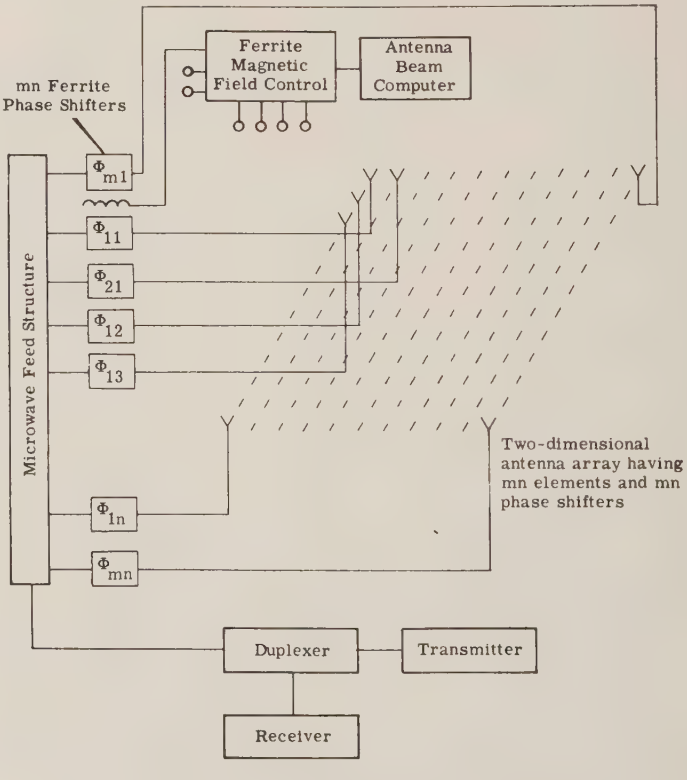


Fig. 6—Phase-phase electronically-scanned radar system.

[3] H. Shnitkin, "Survey of electronically scanned antennas," *Microwave J.*, vol. 3, pp. 67–72; December, 1960.

[4] H. Shnitkin, "Survey of electronically scanned antennas," *Microwave J.*, vol. 4, pp. 57–64; January, 1961.

the system is a three-dimensional phase-phase electronically-scanned radar. In this system, phase shifters are employed in each element, thus making a total of mn phase shifters for the total antenna. The transmitter can operate at fixed frequency, thus removing one of the limitations of the phase-frequency system of Fig. 5. The principal disadvantage of the phase-phase system, however, is the very large number of phase shifters required. Various modifications in the phase-phase system have been developed. Among these are the use of amplifier tubes between the phase shifters and the radiating elements, providing the ability to obtain extremely large total-power output from the system. Other systems make use of heterodyning techniques in which the phase shift is provided at low frequency, thereby eliminating the use of microwave phase-shifting elements, such as ferrites, which are subject to stringent calibration restrictions. Both of the above systems fall into the general category of Class 1, single beam-rapid scan systems.

A number of different antennas have been developed for multiple beam, Class 2 and Class 3, three-dimensional radar systems. Perhaps the oldest and most widely used of these antennas is the so-called stacked-beam antenna, in which several individual feed elements are installed in the vicinity of the focal point of a parabolic reflector. Each of the feeds produces an independent beam whose position in space is a function of the displacement of the feed from the true focus. The number of beams in this type of system is limited as a result of the rapid deterioration of the beam shape as it is displaced from the focal point.

An antenna system which has found wide use in three-dimensional radar is shown in Fig. 7. This makes use of the well-known Luneberg lens. The Luneberg lens is a spherical antenna, composed of dielectric material, with a dielectric constant which is inversely proportional to its position from the center of the sphere. This gives the lens the capability of focusing into a parallel wave front an incident radial wave provided by a radiating element at the opposite end of any diameter through the lens. As shown in Fig. 7, it is therefore possible to install a large number of feeds around the circumference of the lens, each of which produces an independent antenna beam. Individual receivers may be connected to each of the feeds which are, in turn, coupled to a single transmitter, or, as shown in Fig. 7, individual transmitters may be used to provide a total power output equal to nT. The system can be designed as a Class 2 radar by mechanically scanning the feed system in the orthogonal plane, or it can be designed as a Class 3, three-dimensional radar by arranging separate feeds in two dimensions around the circumference of the Luneberg lens.

Another multiple-beam system that has been developed by the Sanders Corporation is shown in Fig. 8. In this system the multiple beams exist in the RECEIVE ONLY mode. A set of tapped delay lines follow individual IF amplifiers which, in turn, are connected to each receiving element of the antenna array. Separate beams are produced by properly connecting the taps of the individual delay lines corresponding to the desired wave front of each beam. The signal is then brought to a high level by providing an amplifier for each of the beams. The principal disadvantages of this type of system are the requirement for individual amplifiers at each receiving element of the array, and the necessity for the maintenance of phase stability in each amplifier. The system must also make use of a separate transmitter which may or may not have multiple beams. The advantage of this type of system is the relatively straightforward means for obtaining a very large number of antenna beams by the use of delay lines which can be readily designed at IF frequencies.

An interesting new antenna system that has been developed at Maxson by Dr. J. Blass is shown in Fig. 9.

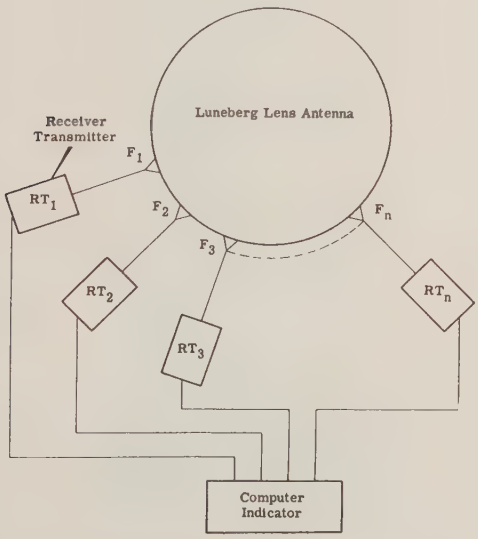


Fig. 7—Multiple-beam Luneberg lens radar system.

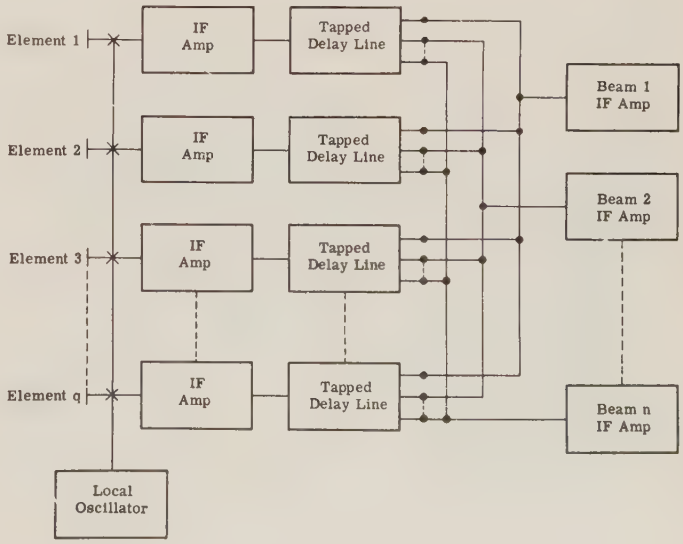


Fig. 8—Delay line multiple-beam radar system.

This antenna, known as the series-fed, multi-directional antenna, is composed of completely passive transmission-line elements connected in matrix form with directional couplers at each of the junction points between the vertical-beam lines and the horizontal-element lines. The direction of each beam is given to the first order by (6) and is determined by the geometric position of each beam line.

$$\sin\theta = \frac{\lambda}{\lambda g}\left[\sec\beta + \tan\beta\right] - \frac{\lambda}{2d}. \qquad (6)$$

Separate receivers and transmitters may be connected to each of the beam line outputs, thus providing either a Class 2 or a Class 3, three-dimensional radar system. The principal advantage of this type of system is the use of completely passive elements which determine beam direction and shape. Thus, the major characteristics of the radar beams will not change as a result of electronic element deterioration. In addition, in this system the number of receivers required is, at most, equal to the number of beams, rather than the number of elements in the antenna. The power-handling capacity of the system is determined by the power-handling capacity of the transmission line used and by the directional couplers. This system lends itself particularly well to applications where long arrays and large numbers of beams are desired.

Another type of linear array, multidirectional antenna arranged in a parallel-feed structure has been designed making use of hybrid elements in a corporate-feed network.

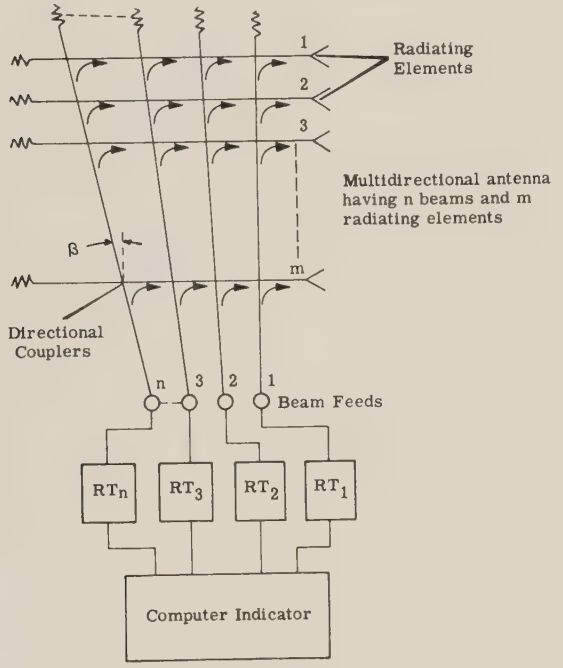


Fig. 9—Series-fed multidirectional antenna radar system.

V. Examples of Three-Dimensional Radar Systems

Two typical examples of three-dimensional radar systems developed at Maxson are given below. The first system is a Class 1, three-dimensional radar similar in design to that shown in Fig. 5, namely, a phase-frequency electronic scan system. An artist's sketch of the complete system is shown in Fig. 10.

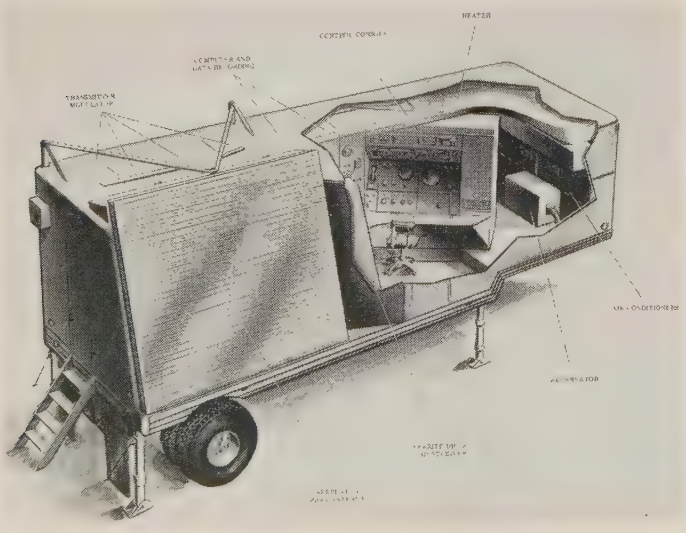


Fig. 10—Class 1, phase-frequency electronic-scan radar system.

Fig. 11—Air-height surveillance radar HRAS-1.

The antenna is seen to be a two-dimensional linear array in which electronic scanning in azimuth is obtained by frequency change, while electronic scan in elevation is provided by change in the magnetic field of ferrite phase-shifter elements. Transmitter, receiver, computer and indicator equipment is contained in the van. The antenna forms part of one side of the van and is permanently fixed to the structure.

An example of an important application of a Class 2, three-dimensional radar system is shown in Fig. 11. This is a photograph of the Maxson Air Height Surveillance Radar system, developed for the Federal Aviation Agency, and installed at their Bureau of Research and Development's facility in Atlantic City, N. J. This system is based on a design shown in Fig. 9. The tower shown in Fig. 11 is 168 feet high and contains a series-fed, multidirectional antenna providing 110 beams in the elevation plane, 0.5° to 40°. A large part of these beams are 3 mils wide at the 3-db point, and provide the capability for resolving two aircraft in the same range /azimuth sector, located 1000 feet apart in altitude at a range of 50 nautical miles, as well as providing a $\pm$500-foot altitude accuracy.

The system shown is passive and obtains its radiation from a standard FAA, ASR radar which provides azimuth and range information. The AHSR equipment provides the third-altitude dimension. The electronics equipment, consisting of 110 receiver channels, beam selector, height computer and three-dimensional indicator, are located in the adjoining electronics building. This system is presently undergoing flight-test evaluation.

VI. Conclusions

The principal approaches used in the design of three-dimensional radar systems have been described. Unfortunately, it has not been possible in this survey to go into great detail regarding the design characteristics of each type of system. It is hoped that the information presented above will provide a good indication of the different approaches that can be used for three-dimensional radar systems. It is expected that, more and more, the modern radar equipments will fall into one of the above categories, and thus provide a substantial increase in performance over existing conventional two-dimensional equipment.

Some of the more important application areas for three-dimensional radars are for use in air-traffic control, artillery and mortar location, missile-range instrumentation, ballistic-missile detection and tracking and satellite surveillance and tracking systems. Many of the systems described above lend themselves particularly to uses in which very large antenna systems are required. In such cases, it is often impractical to consider mechanical rotation in more than a very limited sense for the antenna. Therefore, it is imperative that either electronic scan or multiple-beam systems be provided. Substantial strides can be expected in the coming years in the further development and application of three-dimensional radar systems.

Recent Advancements in Basic Radar Range Calculation Technique*

L. V. BLAKE†, SENIOR MEMBER, IRE

Summary—**A procedure for radar range calculation is described, reflecting current knowledge of the effects of external natural noise sources, atmospheric-absorption losses, and the refractive effect of the normal atmosphere. The range equation is presented in terms of explicitly defined and readily evaluated quantities. Curves and equations are given for evaluating the quantities that are not ordinarily known by direct measurement. Some conventions are proposed for use in general radar range calculation, including an antenna-noise-temperature curve, minimum-detectable signal-to-noise ratio ("visibility factor") curves, and a formula for the reflection coefficient of a rough sea. A noise-temperature table and a work-sheet for range calculation are included in the Appendix.**

Introduction

THE technique of radar maximum range calculation has been refined by taking into account the developments of recent years relative to external natural noise sources, absorption loss in the atmosphere, and normal atmospheric refraction. In the evolution of this technique, the need became apparent for certain standardized assumptions, or conventions, for use in general radar range calculation, especially for military contractual purposes. Some conventions are proposed here for interim use pending possible adoption of official ones by the armed services. Another need is for explicit definition of the range-equation quantities. An attempt is made here to meet that need, and to present the equation in terms of quantities that are readily evaluated, either by direct measurement or by the use of auxiliary equations and curves.

The details of the procedure are presented in terms of conventional surface-based pulse-type monostatic[1] search radar, in the frequency range from about 100 to 10,000 megacycles. However, the calculation method is applicable to radars of other types, usually with minor modification if any.

The range calculated by the method to be described is termed basic because, while it is a starting point for calculations of more complete operational significance, it does not fully take into account some factors that are operationally important, such as clutter echoes, jamming signals, and certain aspects of signal fluctuation.

Here the explanation given of many matters is brief and incomplete, the emphasis being on outlining the method for engineering use rather than on its theoretical and experimental justification. Some of the latter is furnished by the references, and a Naval Research Laboratory report containing detailed explanation is being prepared. An informal NRL report (not generally available) presenting this method of range calculation was issued in November, 1960.

Range Equations

The fundamental radar transmission equation may be found in standard texts such as that of Kerr [8]. For practical calculation, expansion of this equation is required. An excellent practical equation for pulse radars is the one given by Norton and Omberg [4]. This equation is used here with slightly modified notation, and with all the constants and unit-conversion factors gathered into a single numerical factor:

$$R_{50} = 129.2F\left[\frac{P_{t(\mathrm{kw})}\tau_{\mu\mathrm{sec}}G_tG_r\sigma_{50(\mathrm{sq.m.})}}{f_{\mathrm{Mc}}^2T_nV_{o(50)}C_BL}\right]^{1/4}. \qquad (1)$$

The symbols are defined briefly as follows:

R_{50} = range in international nautical miles, 0.5 probability of detection
F = pattern-propagation factor
P_t = transmitter pulse power in kilowatts
$\tau_{\mu\mathrm{sec}}$ = pulse duration in microseconds
G_t, G_r = power gain factors of transmitting and receiving antennas
$\sigma_{50(\mathrm{sq.m.})}$ = median radar cross section of target in square meters
f_{Mc} = radar frequency in megacycles
T_n = effective receiving-system noise temperature in degrees Kelvin
$V_{o(50)}$ = visibility factor for optimum receiver predetection bandwidth and 0.5 probability of detection; ratio of minimum-detectable received pulse energy to noise power per unit bandwidth, both quantities referred to receiver input terminals
C_B = non-optimum-bandwidth correction factor
L = system loss factor.

This equation may also be expressed in a decibel-logarithmic form which is convenient for numerical calculation. Quantities expressed in decibels have the subscript "db." All other quantities are as defined in (1). Logarithms are to the base 10.

$$R_{50} = 100F\cdot\mathrm{antilog}\left[\frac{1}{40}\left\{4.45 + 10\log P_{t(\mathrm{kw})} + 10\log\tau_{\mu\mathrm{sec}} + G_{t(\mathrm{db})} + G_{r(\mathrm{db})} + 10\log\sigma_{50(\mathrm{sq.m.})} - 20\log f_{\mathrm{Mc}} - 10\log T_n - V_{o(50)(\mathrm{db})} - C_{B(\mathrm{db})} - L_{\mathrm{db}}\right\}\right]. \qquad (2)$$

* Received by the PGMIL, January 17, 1961.
† Naval Research Lab., Washington, D.C.
[1] *Monostatic* means that transmitting and receiving antennas are at the same location; a *bistatic* radar is one in which they are fairly widely separated.

The international nautical mile is 1.852 kilometers exactly, or approximately 1.1508 statute miles and 6.0761 thousand feet. If the range is desired in any of these other units, the factor 129.2 in (1) should be multiplied by the appropriate one of these conversion factors (C), and the quantity 4.45 in (2) should be increased by 40 log C.

The statistical aspect of detection imposed by the fluctuating nature of noise and by fluctuations of the target cross-section is recognized by denoting the range for a specified probability of detection and, when appropriate, a specified false-alarm probability. It has become virtually conventional to cite the range for 0.5 probability of detection, here designated R_{50}. The false-alarm probability concept is usually applied only to automatized-detection radars. Following Hall [10], a conventional value of 10^{-10} is suggested. The stated probability of detection applies to a time period equal to the effective integration time of the observer or automatic detection device. The false-alarm probability applies to a time interval approximately equal to N times the reciprocal of the receiver predetection bandwidth, where N is the number of pulses integrated. (Here and in the following, the video bandwidth is assumed to be adequate, *i.e.*, approximately equal to or greater than the predetection bandwidth.) When the effective integration time is less than the scan period of a scanning search radar (as is true of many long-range radars), R_{50} is called the range of 0.5 blip-scan ratio.

Statistical notation is also applied to the target cross-section and to the visibility factor. Although the use of the values σ_{50} and V_{50} does not rigorously insure calculation of exactly R_{50}, the error, if any, is small in most practical cases.

Definition and Evaluation of Range-Equation Factors

The factors P_t, τ, G_t, G_r, σ_{50}, and f_{Mc} may be presumed known from measurement. For some of these quantities, however, there is more than one possible definition. A set of definitions believed to be mutually compatible will be given. Aids to evaluating the factors T_n, $V_{o(50)}$, C_B, L, and F will be presented.

Transmitter Power and Pulse Duration (P_t, τ)

The product $P_t\tau$ represents the pulse energy, which is the time integral of the pulse power envelope, at the transmitter output terminals. The pulse power is

$$P_t = \frac{1}{\tau}\int_{-T/2}^{T/2} W(t)dt, \tag{3}$$

where T is the pulse period and $W(t)$ is the pulse power envelope, excluding any nonuseful portions such as spikes and tails. It is evident that *any* definition of τ will give correct results if the same definition is used in (1) or in (2) and (3). The customary definition is the duration of the pulse between half-power points of the envelope. (This definition is also used in connection with evaluating the bandwidth-correction factor C_B.)

Antenna Gain Factors (G_t, G_r)

The antenna gains G_t and G_r are both defined, on a transmitting basis, as the ratio of the power density radiated in the maximum direction to that, at the same range, of an isotropic antenna radiating the same total power. If the antenna gain is measured in terms of power input rather than power radiated, the resulting figure must be increased by the ratio of the input power to the radiated power to preserve compatibility with the definitions of the loss factor L and the noise temperature T_n.

Radar Cross-Section of Target (σ_{50})

The radar cross-section definition is the standard one given by Kerr [8] and others. The dependence of the cross-section on geometric properties of the target is discussed by Norton and Omberg [4], Ridenour [5], Kerr, and numerous others. As previously stated, σ_{50} denotes the median value when the target cross-section fluctuates. Measured cross-sections are sometimes quoted as the median and sometimes as the mean value, and occasionally other percentile values have been used. It is recommended that the median be adopted as standard. The value one square meter is conventional for assessing the relative range performance of radar systems when no specific target is stipulated. The target fluctuation characteristics (probability distribution, spectrum) are of importance in operational analysis of radar range performance, but are not needed for the basic type of range calculation considered here.

Noise Temperature (T_n)

The receiving-system noise temperature T_n is a fictitious temperature that expresses noise power available at the receiver output as an equivalent available power density at the input terminals of the receiver. This power density is kT_n watts per cycle of bandwidth, where k is 1.38×10^{-23} (Boltzmann's constant). (The total effective input power is this density multiplied by the noise bandwidth of the receiver.) The total noise temperature T_n is the sum of contributions from: 1) external radiating sources, 2) thermal noise due to receiving-transmission-line losses, and 3) internal receiver noise. Each of these three sources is ascribed a noise temperature, termed respectively antenna noise temperature T_a, receiving-transmission-line noise temperature T_r, and effective receiver noise temperature T_e. The temperatures are additive, except that the antenna-noise attenuation by the transmission-line loss factor L_r must be taken into account. Therefore,

$$T_n = T_a/L_r + T_r + T_e. \tag{4}$$

The antenna noise temperature T_a is dependent in a somewhat complicated way on the effective noise temperatures of various radiating sources within the re-

ceiving pattern of the antenna (including sidelobes and backlobes). However, T_a is not directly dependent on the antenna beamwidth and gain. Therefore, it is possible to calculate an antenna temperature which is approximately applicable to any typical radar antenna as a function of frequency. In the microwave region, however, where the thermal noise due to atmospheric absorption is dominant, T_a is also a function of the length of the path in the atmosphere traversed by the beam center, and hence of the beam elevation angle (but not of the target elevation angle, per se).

Curves of antenna temperature are shown in Fig. 1, calculated for the following conditions judged to be typical: 1) average cosmic noise (which actually varies greatly with beam direction, but not in a manner expressible in geocentric coordinates [11]; 2) sun noise temperature 10 times the quiet level, with the sun assumed to be viewed in a sidelobe of unity gain; 3) a cool temperate-zone atmosphere; 4) a contribution of 36°K from ground radiation, which would result (for example) if a ground of blackbody temperature 290°K were viewed over a π-steradian solid angle by sidelobes and back-lobes averaging 0.5 gain (-3 db). This ground-noise contribution, independent of frequency, elevation angle, and beamwidth, is the most arbitrary of the assumptions. It can be justified as a general assumption, but if in a specific case it is not justifiable, the value of T_a given by the curve may be corrected by adding or subtracting an appropriate amount. The dashed curves are for maximum and minimum cosmic and atmospheric noise. Although the solid curves may thus not be correct for every operational condition, they are believed to be suitable as a convention for general range calculation.[2]

The receiving-transmission-line noise temperature T_r is related to the power loss factor L_r and to the thermal (kinetic) temperature of the line T_t by the formula

$$T_r = T_t(1 - 1/L_r). \tag{5}$$

The loss factor L_r represents all available losses preceding the receiver input terminals, including those in the antenna system. It has a multiple role in the range calculation. As indicated by (4), it attenuates the antenna noise, and in accordance with (5) it results in generation of thermal noise. It also attenuates the radar echo signal. The definitions given for G_r, T_a, and T_r are designed to allow a single definition of L_r in these three roles.

If various lossy components of the transmission-line system operate at appreciably different thermal temperatures, T_r must be computed from a cascade formula. A suggested conventional value for T_t is 290°K, applicable except when the transmission line or some of its components may be expected to operate at considerably above or below ambient temperature.

[2] A Naval Research Lab. report discussing the details of the assumptions and method of calculating these curves is in preparation. Also, Hansen [15] and Hogg and Mumford [17] have published excellent reviews on antenna-temperature calculation.

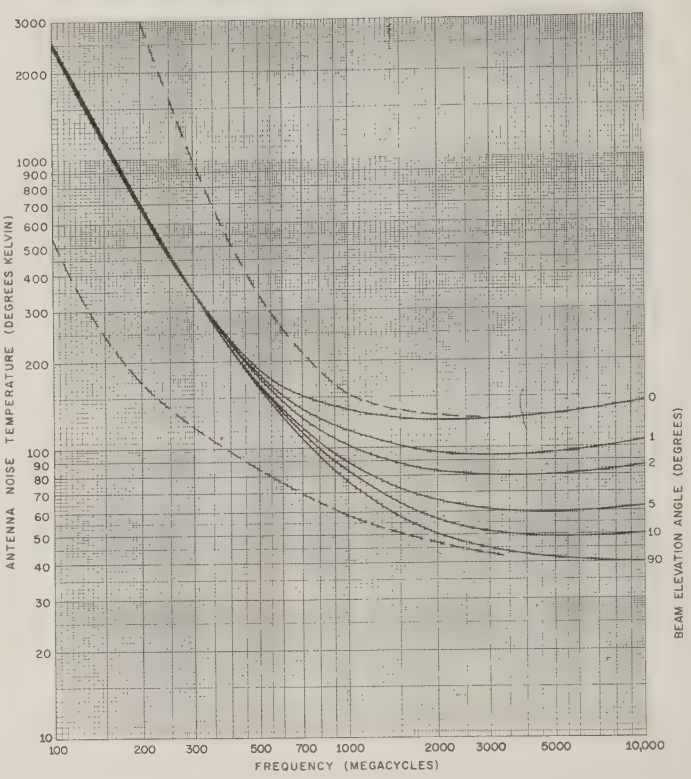


Fig. 1—Antenna temperature for typical conditions. Dashed curves are for maximum and minimum cosmic and atmospheric noise.

The effective receiver noise temperature T_e is related to the receiver noise factor, $\overline{NF}$ (IRE Standard 53 IRE 17.S1),[3] by the formula

$$T_e = (\overline{NF} - 1)T_o, \tag{6}$$

where $T_o = 290$°K, the reference temperature for noise-factor measurement.

Visibility Factor ($V_{o(50)}$) and Bandwidth Correction Factor (C_B)

The visibility factor is a specialized form of the minimum detectable signal-to-noise ratio, for specified probability of detection. It is expressed here as the product of the optimum-bandwidth value $V_{o(50)}$ and a bandwidth-correction factor C_B. It is a function of the number of pulses integrated. Values for the case of a human observer of a cathode-ray-tube indicator, based on World War II experiments at the M.I.T. Radiation Laboratory [3], [7], are given by Figs. 2 and 3. The equation for the number of pulses integrated by a non-scanning radar is

$$N = \overline{PRF} \cdot t_i, \tag{7}$$

where $\overline{PRF}$ is the radar pulse rate and t_i is the effective integration time. The characteristics of the integrator determine t_i. Its value for electronic storage or delay devices can usually be assessed readily, but it is difficult to assign a numerical value for the combination of a human

[3] "Standards on receivers: Definitions of terms, 1952," Proc. IRE, vol. 40, pp. 1681–1685; December, 1952.

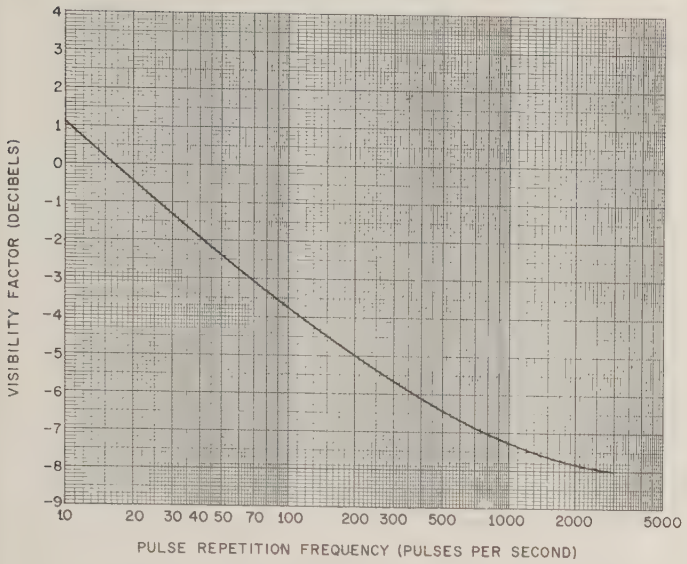


Fig. 2—Visibility factor $V_{o(50)db}$ for type-A cathode-ray-tube display, based on M.I.T. Radiation Lab. (pre-1946) experiments [3], [7].

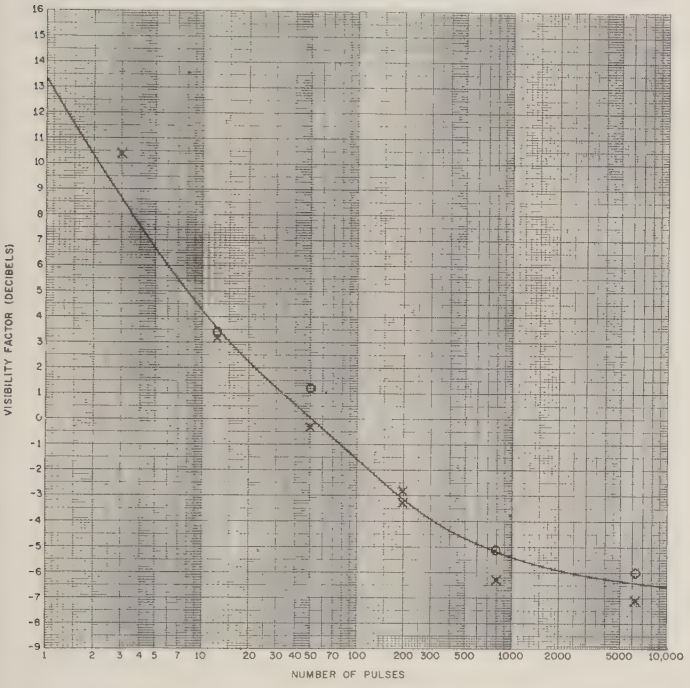


Fig. 3—Visibility factor $V_{o(50)db}$ for PPI cathode-ray-tube display, based on M.I.T. Radiation Lab. (pre-1946) experiments [3], [7].

observer and cathode-ray tube. This problem is bypassed in Fig. 2 by employing the $\overline{P}\,\overline{R}\overline{F}$ directly as the variable. Fig. 3 may be used for scanning radars without exact knowledge of t_i when it can be assumed that t_i is shorter than the scan period and longer than the interval required for the beam to traverse the target. In this case the number of pulses integrated is taken to be the number occurring while the target is within the half-power limits of the (one-way) antenna pattern during a single scan.

M.I.T. Radiation Laboratory experiments during World War II indicate [3], [7] that t_i may be as great as 6 to 10 seconds for highly trained observers, but many experienced radar engineers feel that a somewhat shorter time is probably characteristic of the average observer. It is suggested that a more conservative value, $t_i = 2$ seconds, be assumed for conventional range-calculation purposes. Thus, for example, if a radar scans in azimuth at a rate faster than 30 revolutions per minute ($\overline{R}\overline{P}\overline{M}$), some scan-to-scan integration should be assumed, and the number of pulses integrated would be the number occurring per beamwidth per scan multiplied by $\overline{R}\overline{P}\overline{M}/30$.

The number of pulses thus computed may be used in connection with Fig. 3 to determine a value of $V_{o(50)}$, in decibels, applicable to (3) when a PPI (or any similar intensity-modulated cathode-ray-tube display) is used with a scanning radar and a human observer, without other integrating devices. (Generally, supplementary integrators improve the visiblity factor only if they have a longer effective integration time than that of the human observer, or if they operate as predetection integrators.) The leveling-off of the curves at very low values of $V_{o(50)}$ is attributed to the limited contrast discernibility of the human eye-brain combination [3], [7].

Figs. 2 and 3 represent, implicitly, some unstated value of false-alarm probability. Therefore, these two curves are proposed as conventions (rather than as precise values) for use in calculating the range of radars that use cathode-ray-tube indicators and human observers. Experience with these curves indicates that the calculated ranges are in reasonable agreement with experimental results.

The bandwidth correction factor has been determined as an empirical formula by Haeff and others [2] and in the form of curves by Lawson and Uhlenbeck [3], [7]. These empirical results differ from results of purely theoretical analysis, presumably [7] because of certain characteristics of the human eye-brain combination. Haeff's formula (which is in general agreement with the Lawson-Uhlenbeck results), modified to reflect the finding that optimum bandwidth is $1.2/\tau$, is

$$C_B = \frac{B\tau}{4.8}\left(1 + \frac{1.2}{B\tau}\right)^2. \tag{8}$$

Decibel values of C_B based on this formula are given by Fig. 4 in terms of the ratio of the actual bandwidth B to the optimum value, $B_{opt} = 1.2/\tau$. This assessment of the optimum value [7] is based on definition of bandwidth and pulse length as the half-power values, and on the conditions that the pulse is approximately rectangular and that the receiver has a conventional double-tuned IF amplifier, or one with similar pass band shape. It is further restricted to the case of detection by a human observer of a conventional cathode-ray-tube indicator.

For automatized radar detection curves of minimum-detectable signal-to-noise *power ratio*, defined at the

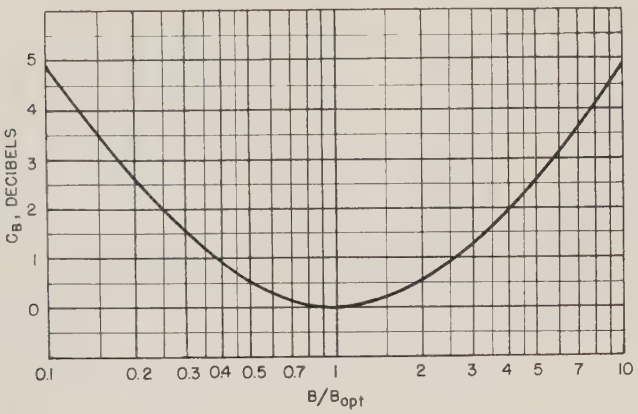


Fig. 4—Empirical curve for bandwidth-correction factor as a function of ratio of actual predetection bandwidth B to optimum value $B_{opt}=1.2/\tau$.

detector input terminals, may be calculated, as exemplified by the work of Marcum [18]. Curves of this type calculated by the author, following methods described by North [1], are shown in Fig. 5. They are for a fixed-threshold-level decision-making device, preceded by a linear-rectifier detector and a perfect-memory linear video integrator. The N pulses integrated are assumed to be of constant amplitude. The quantity represented by these curves may be called detectability factor D_{50} to distinguish it from the visibility factor. D_{50} is a power ratio, defined at the detector terminals rather than at the receiver input terminals. The selective circuits of the receiver intervene. North [1] has shown that if the receiver pass band transfer characteristic "looks like the conjugate of the spectrum of the echo at the antenna," $V_{o(50)}$ and D_{50} are equal, but otherwise they are not. In general,

$$V_{o(50)} = mD_{50}, \tag{9}$$

where $m \geqq 1$ may be called a matching factor. Thus the curves of Fig. 5 may be used with (1) by applying (9). The values of C_B given by (8) and Fig. 4 do not apply in general for automatized detection. Lawson and Uhlenbeck [7][4] have analyzed these matters for certain specific cases (though not in the terminology employed here).

If perfect predetection integration were assumed, each curve would have the same value shown in Fig. 5 for $N=1$, denoted $D_{50(db)}(1)$ but would follow the law

$$D_{50(db)}(N) = D_{50(db)}(1) - 10 \log N. \tag{10}$$

Loss Factors

Loss factor is defined as the ratio of the power input to the power output of the lossy element of the system (*i.e.*, reciprocal of gain). The general loss factor L is the product of numerous specific loss factors, certain ones of

[4] Reference [7], pp. 204–210.

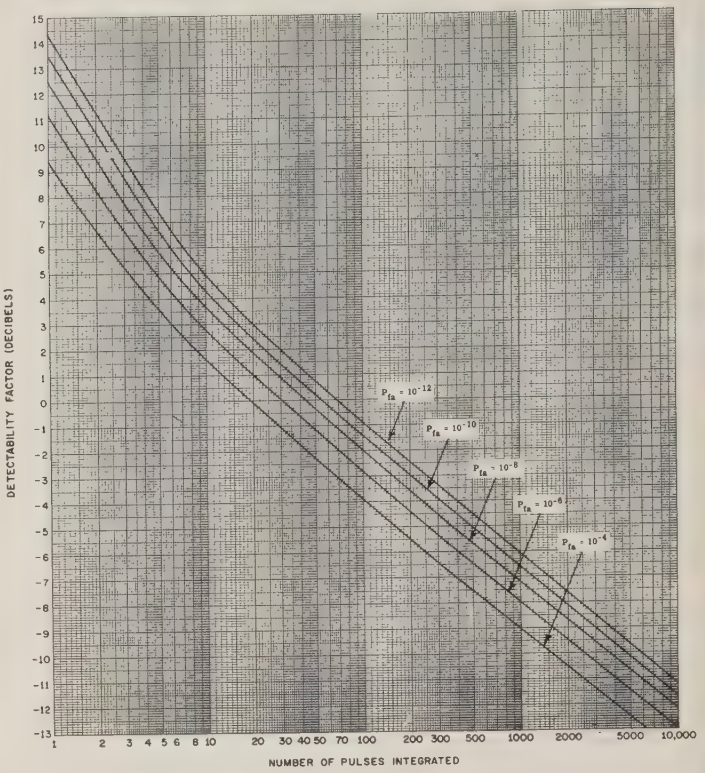


Fig. 5—Detectability factor $D_{50(db)}$ calculated for a linear-rectifier detector, followed by a perfect-memory linear video integrator and a fixed-threshold-level automatized-detection device, for several values of false-alarm probability P_{fa}.

which are generally present. (L_{db} is of course the sum of the component loss factors expressed in decibels.)

The factor L_t accounts for the total transmission-line loss that occurs during transmitting, between the transmitter output terminals and radiating surfaces of the antenna. The factor L_r is the transmission-line loss on reception, analogously defined (but not necessarily equal to L_t). However, L_r is the so-called available loss,[5] or ratio of available power at the antenna to available power at the receiver input terminals, whereas L_t is the actual loss (ratio of actual transmitter output power to power radiated by the antenna). Available power is that which would be delivered to a matched load impedance (complex conjugate of source impedance).

L_p, the antenna-pattern loss factor, accounts for the fact that the gain of a scanning antenna, in the target direction, varies from pulse to pulse in accordance with the antenna pattern, while the antenna gain factors in the equation are applicable to targets in the beam maximum. It also takes into account the arbitrary designation of beamwidth, for the purpose of counting the number of pulses integrated, as the half-power value. Anal-

[5] The necessity for this definition of L_r was called to the author's attention by L. E. Davies of Stanford Research Institute.

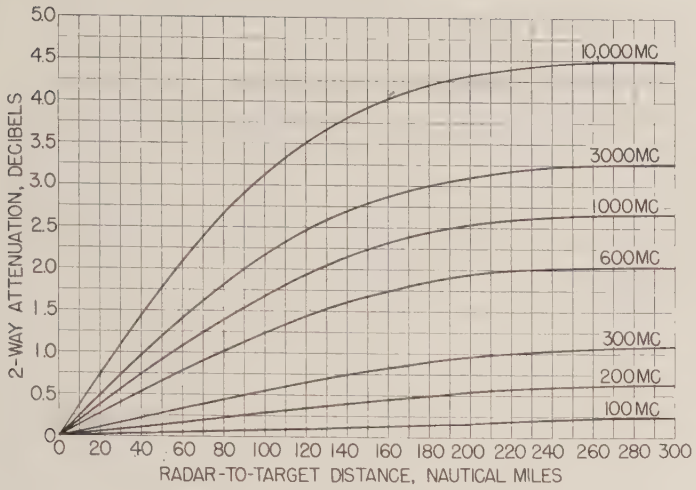


Fig. 6—Atmospheric-absorption loss $L_{a(db)}$, calculated for 0° target elevation angle.

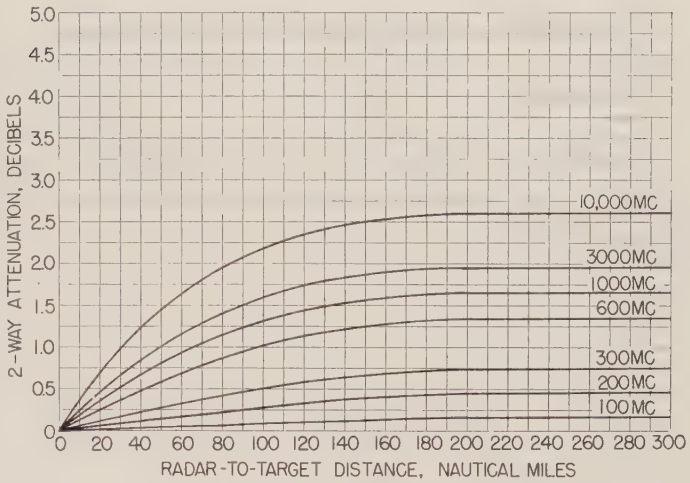


Fig. 7—Atmospheric-absorption loss $L_{a(db)}$ calculated for 1° target elevation angle.

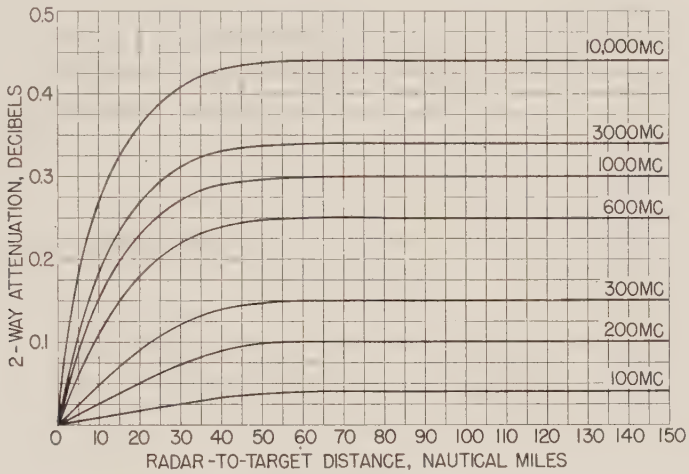


Fig. 8—Atmospheric-absorption loss $L_{a(db)}$ calculated for 10° target elevation angle.

ysis of these matters [9] indicates that a loss factor of 1.45 (1.6 db) is appropriate in the case of an azimuth-scanning radar. For bidirectionally scanning radar, this factor is approximately the square of the loss factor for a unidirectional scan—*i.e.*, 2.1 or 3.2 db. For a non-scanning radar, $L_p = 1$ (0 db). (If the target is not in the beam maximum for such a radar, appropriate correction should be made in the pattern-propagation factor F. In the case of a unidirectional scanning radar, if the target is displaced from the beam maximum in the direction orthogonal to the scanning direction, this should also be taken into account by the pattern-propagation factor.)

The factor L_a is for loss due to absorption in the propagation medium. Curves for this loss in the atmosphere as a function of target range, for certain elevation angles, are given by Figs. 6 to 8. They were calculated [14] on the basis of Van Vleck's theory of absorption by oxygen and water vapor. ([14] contains curves for additional elevation angles.) Since this loss depends on the range, it is necessary first to calculate the range for no loss ($L_a = 1$), and then apply a correction based on the loss factor determined for this range from curves of the type of Figs. 6 to 8. Further correction may be made, if necessary, on the basis of a revised loss factor for the corrected range.

Additional losses may occur in special cases. Among the possibilities are collapsing loss, video mixing loss, sweep-speed loss, and pulse-length loss (due to finite excitation time of certain types of array antennas). The basis of calculation of some of these losses is given by Lawson and Uhlenbeck [7]. A quantity sometimes called operator loss is already included in Figs. 2 and 3. Inclusion of a so-called system-degradation loss to account for poor system maintenance and adjustment is deprecated.

Pattern-Propagation Factor

The pattern-propagation factor F, as defined by Kerr [8], is the ratio of the actual field strength at the target to that which would be observed in free space at the same range, in the beam maximum. Here the further provision is made that the field strengths in this definition shall be those that would exist in the absence of any propagation-medium absorption. Formulas for computing F are given by Kerr.

A troublesome factor in evaluating F in the case of sea-reflection has been determination of the magnitude of the reflection coefficient ρ. The value for a smooth sea ρ_o is given by Kerr[6] as a function of the polarization, frequency, and grazing angle. For a rough sea, the following empirical formula has been devised:

$$\rho = \rho_o \left[\frac{k}{k + A} \right]. \tag{11}$$

For a 6-foot wave height, calculated values conform approximately to experience and to Rayleigh's criterion when $A = f_{Mc}^2 (\sin^2 \theta) \exp [-0.05(90° - \theta°)]$ and $k = 25$, where f_{Mc} is the frequency in megacycles and $\theta°$ is the grazing angle in degrees. (For low antenna height, θ may be taken as the target elevation angle.)

This is not a formula based on rigorous analysis or experiments. It is offered primarily as a convention for use until better theoretical or experimental information is available; it gives "reasonable" values.

Vertical Coverage Diagrams

Radar maximum range is generally a function of the *target elevation angle* θ, and can be described adequately only in terms of a vertical coverage diagram. It has until recently been customary to compute the range-height-angle relationship assuming a *linear* decay of refractive index with height, but this assumption leads to serious errors at low angles and long ranges. A better representation of the actual refractive-index variation, as shown by Bauer [12] and Bean and Thayer [13], is a negative-exponential model. For general use, the following particularization of the Bureau of Standards CRPL Exponential Reference Atmosphere described by Bean and Thayer is suggested:

$$n(h) = 1 + 0.000313 \exp (-0.04385h), \tag{12}$$

where n is the refractive index and h is the height in thousands of feet. Bean[7] states that this surface value of n (1.000313) was obtained by averaging about 2×10^6 observations for a period of eight years over the United States. Fig. 9 is a plot of range-height-angle values calculated (using the Naval Research Laboratory NAREC Computer) for this model. Bean and Thayer [13], in NBS Monograph 4, have published tables of values (expressed in kilometers and milliradians) for this and other models. Bauer [12] has also published values for an exponential model. Fig. 9 is designed to present ray paths as straight lines, a feature that facilitates the plotting of coverage diagrams.

Appendix

Aids to Radar Range Calculation

A work-sheet form has been devised for use with the curves and equations of this paper and with Tables I and II (see page 163). This form, shown following Table II, greatly facilitates range calculations and

[6] Reference [8], pp. 402–403.

[7] B. R. Bean, Natl. Bureau of Standards, Boulder Labs., Boulder, Colo., in a letter to the author; December 8, 1960.

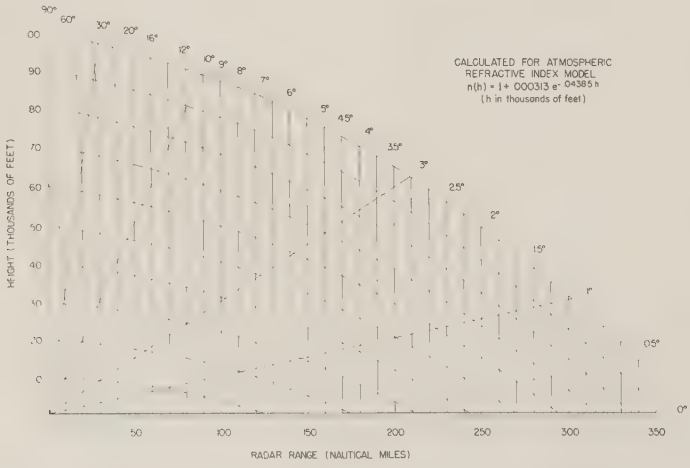


Fig. 9—Radar range-height-angle chart calculated for an exponential model of the atmospheric refractive index. ("Radar range" means "distance along the ray path." Elevation angles are angles of rays with respect to horizontal at radar antenna.)

allows inexperienced personnel to perform them. It is based on (2) and is largely self-explanatory. The quantity L_x denotes the aggregate of any losses that may occur in special cases, in addition to the ones specifically provided for. Values of $L_{a(db)}$ for elevation angles or frequencies not shown in Figs. 6 to 8 may be estimated by interpolation. Table I, together with Fig. 1, is useful for computing the system noise temperature.

The following formulas are useful for computing the number of pulses N illuminating the target per azimuth scan of a scanning radar, as required for use with Figs. 3 and 5. For azimuth-only scanning,

$$N = \frac{\theta_h \cdot \overline{PRF}}{6 \cdot \cos\theta \cdot \overline{RPM}}, \tag{13}$$

where θ_h is the horizontal beamwidth in degrees (the value applicable at the target elevation angle), $\overline{PRF}$ is the pulse repetition frequency in pulses per second, $\overline{RPM}$ is the scanning speed in revolutions per minute, and θ is the target elevation angle. For an azimuth-and-elevation-scanning radar (assuming that there are many elevation scans per azimuth scan) the formula is

$$N = \frac{\theta_h \theta_v \overline{PRF}}{6(\cos\theta)\omega_v t_v \overline{RPM}}, \tag{14}$$

where θ_v is the vertical beamwidth in degrees, ω_v is the vertical scanning speed in degrees per second (at the elevation angle of the target), and t_v is the vertical scanning period in seconds (including dead time, if any). These formulas apply as long as $\theta_h/\cos\theta$ is not greater than about 90°. For greater values, more complicated formulas are required.

TABLE I

DECIBELS, POWER RATIOS, AND EFFECTIVE NOISE TEMPERATURES

If the second column represents a loss ratio, L, the third column is its reciprocal, which occurs in the formula for the effective noise temperature of a lossy receiving transmission line or propagation medium:

$$T_r = T_t\left(1 - \frac{1}{L}\right)$$

where T_r is the effective noise temperature and T_t is the thermal temperature (Kelvin) of the line or medium. The fourth column is the resulting noise temperature, calculated from this equation, assuming $T_t = 290°K$. If in the actual case T_t has some other value, multiply the values of T_r by $T_t/290$.

The second column also represents the receiver noise factor, $\overline{NF}$, expressed as a power ratio, corresponding to decibel noise-factor values represented by the first column. The fifth column is the corresponding effective receiver noise temperature, $T_e = (\overline{NF} - 1)T_o$. ($T_o = 290°K$).

Temperature conversion relations: $T_{Kelvin} = 273.16 + T_{Centigrade} = 255.38 + (5/9)\,T_{Fahrenheit}$

Decibels ±	Power Ratios + (NF)	Power Ratios - (1/L_r)	T_r °Kelvin	T_e °Kelvin
0	1.0000	1.0000	0.00	0.00
0.01	1.0023	.9977	0.67	0.67
0.02	1.0046	.9954	1.33	1.33
0.03	1.0069	.9931	2.00	2.00
0.04	1.0093	.9908	2.67	2.70
0.05	1.0116	.9886	3.31	3.36
0.06	1.0139	.9863	3.97	4.03
0.07	1.0162	.9840	4.64	4.70
0.08	1.0186	.9818	5.28	5.39
0.09	1.0209	.9795	5.95	6.06
0.10	1.0233	.9772	6.61	6.76
0.15	1.0351	.9661	9.83	10.2
0.20	1.0471	.9550	13.1	13.7
0.25	1.0593	.9441	16.2	17.2
0.30	1.0715	.9333	19.3	20.7
0.35	1.0839	.9226	22.4	24.3
0.40	1.0965	.9120	25.5	28.0
0.45	1.1092	.9016	28.5	31.7
0.50	1.1220	.8913	31.5	35.4
0.55	1.1350	.8810	34.5	39.2
0.60	1.1482	.8710	37.4	43.0
0.65	1.1614	.8610	40.3	46.8
0.70	1.1749	.8511	43.2	50.7
0.75	1.1885	.8414	46.0	54.7
0.80	1.2023	.8318	48.8	58.7
0.85	1.2162	.8222	51.6	62.7
0.90	1.2303	.8128	54.3	66.8
0.95	1.2445	.8035	57.0	70.9
1.00	1.2589	.7943	59.7	75.1
1.1	1.288	.7762	64.9	83.5
1.2	1.318	.7586	70.0	92.2
1.3	1.349	.7413	75.0	101
1.4	1.380	.7244	79.9	110
1.5	1.413	.7079	84.7	120
1.6	1.445	.6918	89.4	129
1.7	1.479	.6761	93.9	139
1.8	1.514	.6607	98.4	149
1.9	1.549	.6457	103	159
2.0	1.585	.6310	107	170
2.1	1.622	.6166	111	180
2.2	1.660	.6026	115	191
2.3	1.698	.5888	119	202
2.4	1.738	.5754	123	214
2.5	1.778	.5623	127	226
2.6	1.820	.5495	131	238
2.7	1.862	.5370	134	250
2.8	1.905	.5248	138	262
2.9	1.950	.5129	141	276
3.0	1.995	.5012	145	289
3.1	2.042	.4898	148	302
3.2	2.089	.4786	151	316
3.3	2.138	.4677	154	330
3.4	2.188	.4571	157	345
3.5	2.239	.4467	160	359
3.6	2.291	.4365	163	374
3.7	2.344	.4266	166	390
3.8	2.399	.4169	169	406
3.9	2.455	.4074	172	422
4.0	2.512	.3981	175	438
4.1	2.570	.3890	177	455
4.2	2.630	.3802	180	473
4.3	2.692	.3715	182	491
4.4	2.754	.3631	185	509
4.5	2.818	.3548	187	527
4.6	2.884	.3467	189	546
4.7	2.951	.3388	192	566
4.8	3.020	.3311	194	586
4.9	3.090	.3236	196	606
5.0	3.162	.3162	198	627
5.1	3.236	.3090	200	648
5.2	3.311	.3020	202	670
5.3	3.388	.2951	204	693
5.4	3.467	.2884	206	715
5.5	3.548	.2818	208	739
5.6	3.631	.2754	210	763
5.7	3.715	.2692	212	787
5.8	3.802	.2630	214	813
5.9	3.890	.2570	215	838
6.0	3.981	.2512	217	864
6.1	4.074	.2455	219	891
6.2	4.169	.2399	220	919
6.3	4.266	.2344	222	947
6.4	4.365	.2291	224	976
6.5	4.467	.2239	225	1005
6.6	4.571	.2188	227	1036
6.7	4.677	.2138	228	1066
6.8	4.786	.2089	229	1098
6.9	4.898	.2042	231	1130
7.0	5.012	.1995	232	1163
7.1	5.129	.1950	233	1197
7.2	5.248	.1905	235	1232
7.3	5.370	.1862	236	1267
7.4	5.495	.1820	237	1304
7.5	5.623	.1778	238	1341
7.6	5.754	.1738	240	1379
7.7	5.888	.1698	241	1418
7.8	6.026	.1660	242	1458
7.9	6.166	.1622	243	1498
8.0	6.310	.1585	244	1540
8.1	6.457	.1549	245	1583
8.2	6.607	.1514	246	1626
8.3	6.761	.1479	247	1671
8.4	6.918	.1445	248	1716
8.5	7.079	.1413	249	1763
8.6	7.244	.1380	250	1811
8.7	7.413	.1349	251	1860
8.8	7.586	.1318	252	1910
8.9	7.762	.1288	253	1961
9.0	7.943	.1259	253	2013
9.1	8.128	.1230	254	2067
9.2	8.318	.1202	255	2122
9.3	8.511	.1175	256	2178
9.4	8.710	.1148	257	2236
9.5	8.913	.1122	257	2295
9.6	9.120	.1096	258	2355
9.7	9.333	.1072	259	2417
9.8	9.550	.1047	260	2480
9.9	9.772	.1023	260	2544
10.0	10.000	.1000	261	2610

TABLE II

Radar Range Factors for System Power Change from 0 to 40 Decibels

(In steps of 0.1 decibel)

The table is intended for use with an equation of the type:

$$R = k\left[\frac{P_t G^2 \lambda^2 \sigma F^4}{P_r L}\right]^{1/4} = k P^{1/4}; \text{ i.e., } R \propto P^{1/4}$$

where R is the radar range and P may be regarded as an equivalent system power variable. The table is based on the relation:

$$R/R_o = \text{antilog}\left[\frac{1}{40}(10 \log P/P_o)\right]$$

where R/R_o is the range factor, and $10 \log P/P_o$ is the power change in decibels. P_t is transmitter power, G antenna gain, λ wavelength, σ target cross section, L loss factor, F pattern-propagation factor, and P_r received echo power.

Range factors for power changes greater than 40 db can be obtained from the table by the following procedure: (1) Subtract from the absolute value of the power change in db the integral multiple of 40 which results in a positive remainder less than 40; (2) Look up the range factor corresponding to the remainder; (3) Shift the decimal point one place for each 40 db subtracted; for range increase, shift to the right, for decrease shift to the left. For example, the range increase for a power change of 47.3 db is 15.22, and for 87.3 it is 152.2, because for 7.3 db it is 1.522. The decrease factor for 47.3 db is 0.06569, and for 87.3 it is 0.006569, etc.

Power Change, Decibels ±	Range Increase Factor (Decimal Point)	Range Decrease Factor (Decimal Point)	± Decibels Power Change
0.0	1 0000	1 0 000	40.0
0.1	1 0058	9 943	39.9
0.2	1 0116	9 886	39.8
0.3	1 0174	9 829	39.7
0.4	1 0233	9 772	39.6
0.5	1 0292	9 716	39.5
0.6	1 0351	9 661	39.4
0.7	1 0411	9 605	39.3
0.8	1 0471	9 550	39.2
0.9	1 0532	9 495	39.1
1.0	1 0593	9 441	39.0
1.1	1 065	9 386	38.9
1.2	1 072	9 333	38.8
1.3	1 078	9 279	38.7
1.4	1 084	9 226	38.6
1.5	1 090	9 173	38.5
1.6	1 096	9 120	38.4
1.7	1 103	9 068	38.3
1.8	1 109	9 016	38.2
1.9	1 116	8 964	38.1
2.0	1 122	8 913	38.0
2.1	1 129	8 861	37.9
2.2	1 135	8 810	37.8
2.3	1 142	8 760	37.7
2.4	1 148	8 710	37.6
2.5	1 155	8 660	37.5
2.6	1 161	8 610	37.4
2.7	1 168	8 561	37.3
2.8	1 175	8 511	37.2
2.9	1 182	8 463	37.1
3.0	1 189	8 414	37.0
3.1	1 195	8 366	36.9
3.2	1 202	8 318	36.8
3.3	1 209	8 270	36.7
3.4	1 216	8 222	36.6
3.5	1 223	8 175	36.5
3.6	1 230	8 128	36.4
3.7	1 237	8 082	36.3
3.8	1 245	8 035	36.2
3.9	1 252	7 989	36.1
4.0	1 259	7 943	36.0
4.1	1 266	7 898	35.9
4.2	1 274	7 852	35.8
4.3	1 281	7 807	35.7
4.4	1 288	7 763	35.6
4.5	1 296	7 718	35.5
4.6	1 303	7 674	35.4
4.7	1 311	7 630	35.3
4.8	1 318	7 586	35.2
4.9	1 326	7 542	35.1
	Decimal Point Range Decrease Factor	Decimal Point Range Increase Factor	± Decibels Power Change

Power Change, Decibels ±	Range Increase Factor (Decimal Point)	Range Decrease Factor (Decimal Point)	± Decibels Power Change
5.0	1 334	7 499	35.0
5.1	1 341	7 456	34.9
5.2	1 349	7 413	34.8
5.3	1 357	7 371	34.7
5.4	1 365	7 328	34.6
5.5	1 372	7 286	34.5
5.6	1 380	7 244	34.4
5.7	1 388	7 203	34.3
5.8	1 396	7 162	34.2
5.9	1 404	7 120	34.1
6.0	1 413	7 080	34.0
6.1	1 421	7 039	33.9
6.2	1 429	6 998	33.8
6.3	1 437	6 958	33.7
6.4	1 445	6 918	33.6
6.5	1 454	6 879	33.5
6.6	1 462	6 839	33.4
6.7	1 471	6 800	33.3
6.8	1 479	6 761	33.2
6.9	1 488	6 722	33.1
7.0	1 496	6 683	33.0
7.1	1 505	6 645	32.9
7.2	1 514	6 607	32.8
7.3	1 522	6 569	32.7
7.4	1 531	6 531	32.6
7.5	1 540	6 494	32.5
7.6	1 549	6 457	32.4
7.7	1 558	6 420	32.3
7.8	1 567	6 383	32.2
7.9	1 576	6 346	32.1
8.0	1 585	6 310	32.0
8.1	1 594	6 273	31.9
8.2	1 603	6 237	31.8
8.3	1 612	6 202	31.7
8.4	1 622	6 166	31.6
8.5	1 631	6 131	31.5
8.6	1 641	6 095	31.4
8.7	1 650	6 061	31.3
8.8	1 660	6 026	31.2
8.9	1 669	5 991	31.1
9.0	1 679	5 957	31.0
9.1	1 689	5 923	30.9
9.2	1 698	5 888	30.8
9.3	1 708	5 855	30.7
9.4	1 718	5 821	30.6
9.5	1 728	5 788	30.5
9.6	1 738	5 754	30.4
9.7	1 748	5 721	30.3
9.8	1 758	5 689	30.2
9.9	1 768	5 656	30.1
	Decimal Point Range Decrease Factor	Decimal Point Range Increase Factor	± Decibels Power Change

Power Change, Decibels ±	Range Increase Factor (Decimal Point)	Range Decrease Factor (Decimal Point)	± Decibels Power Change
10.0	1 778	5 623	30.0
10.1	1 789	5 591	29.9
10.2	1 799	5 559	29.8
10.3	1 809	5 527	29.7
10.4	1 819	5 495	29.6
10.5	1 830	5 464	29.5
10.6	1 841	5 433	29.4
10.7	1 851	5 401	29.3
10.8	1 862	5 370	29.2
10.9	1 873	5 340	29.1
11.0	1 884	5 309	29.0
11.1	1 895	5 278	28.9
11.2	1 905	5 248	28.8
11.3	1 916	5 218	28.7
11.4	1 928	5 188	28.6
11.5	1 939	5 158	28.5
11.6	1 950	5 129	28.4
11.7	1 961	5 099	28.3
11.8	1 972	5 070	28.2
11.9	1 984	5 041	28.1
12.0	1 995	5 012	28.0
12.1	2 007	4 983	27.9
12.2	2 018	4 954	27.8
12.3	2 030	4 926	27.7
12.4	2 042	4 898	27.6
12.5	2 054	4 870	27.5
12.6	2 065	4 842	27.4
12.7	2 077	4 814	27.3
12.8	2 089	4 786	27.2
12.9	2 101	4 759	27.1
13.0	2 113	4 732	27.0
13.1	2 126	4 704	26.9
13.2	2 138	4 677	26.8
13.3	2 150	4 650	26.7
13.4	2 163	4 624	26.6
13.5	2 175	4 597	26.5
13.6	2 188	4 571	26.4
13.7	2 200	4 545	26.3
13.8	2 213	4 519	26.2
13.9	2 226	4 493	26.1
14.0	2 239	4 467	26.0
14.1	2 252	4 441	25.9
14.2	2 265	4 416	25.8
14.3	2 278	4 390	25.7
14.4	2 291	4 365	25.6
14.5	2 304	4 340	25.5
14.6	2 317	4 315	25.4
14.7	2 331	4 290	25.3
14.8	2 344	4 266	25.2
14.9	2 358	4 241	25.1
	Decimal Point Range Decrease Factor	Decimal Point Range Increase Factor	± Decibels Power Change

Power Change, Decibels ±	Range Increase Factor (Decimal Point)	Range Decrease Factor (Decimal Point)	± Decibels Power Change
15.0	2 371	4 217	25.0
15.1	2 385	4 193	24.9
15.2	2 399	4 169	24.8
15.3	2 413	4 145	24.7
15.4	2 427	4 121	24.6
15.5	2 441	4 097	24.5
15.6	2 455	4 074	24.4
15.7	2 469	4 050	24.3
15.8	2 483	4 027	24.2
15.9	2 497	4 004	24.1
16.0	2 512	3 981	24.0
16.1	2 526	3 958	23.9
16.2	2 541	3 936	23.8
16.3	2 556	3 913	23.7
16.4	2 570	3 890	23.6
16.5	2 585	3 868	23.5
16.6	2 600	3 846	23.4
16.7	2 615	3 824	23.3
16.8	2 630	3 802	23.2
16.9	2 645	3 780	23.1
17.0	2 661	3 758	23.0
17.1	2 676	3 737	22.9
17.2	2 692	3 715	22.8
17.3	2 707	3 694	22.7
17.4	2 723	3 673	22.6
17.5	2 738	3 652	22.5
17.6	2 754	3 631	22.4
17.7	2 770	3 610	22.3
17.8	2 786	3 589	22.2
17.9	2 802	3 569	22.1
18.0	2 818	3 548	22.0
18.1	2 835	3 528	21.9
18.2	2 851	3 508	21.8
18.3	2 867	3 487	21.7
18.4	2 884	3 467	21.6
18.5	2 901	3 447	21.5
18.6	2 917	3 428	21.4
18.7	2 934	3 408	21.3
18.8	2 951	3 388	21.2
18.9	2 968	3 369	21.1
19.0	2 985	3 350	21.0
19.1	3 003	3 330	20.9
19.2	3 020	3 311	20.8
19.3	3 037	3 292	20.7
19.4	3 055	3 273	20.6
19.5	3 073	3 255	20.5
19.6	3 090	3 236	20.4
19.7	3 108	3 217	20.3
19.8	3 126	3 199	20.2
19.9	3 144	3 181	20.1
20.0	3 162	3 162	20.0
	Decimal Point Range Decrease Factor	Decimal Point Range Increase Factor	± Decibels Power Change

PULSE-RADAR RANGE-CALCULATION WORK SHEET

1. Compute system noise temperature, T_n, following outline in section (1) below.

2. Enter range factors known in other than decibel form in section (2) below, for reference.

3. Enter logarithmic and decibel values in section (3) below, positive values in plus column, negative in minus. (Example: If $V_{o(50)(db)}$ as given by Fig. 2 or 3 is negative, then $-V_{o(50)(db)}$ is positive, goes in plus column.) To convert range factors to decibel values, use Table 1. For $C_{B(db)}$ use Fig. 4.

Radar antenna height: h = ft. Target elevation angle: θ = °. (See Fig. 9.)

(1) COMPUTATION OF T_n:

$T_n = T_a/L_r + T_r + T_e$

(a) For general range computation, use Figure 1 for T_a.

(b) To find $1/L_r$, given $L_{r(db)}$, use first and third columns of Table 1.

(c) Also in Table 1, opposite $L_{r(db)}$ in first column, read T_r in fourth column.

Note: If thermal temperature (T_t) of transmission line is appreciably different from 290°K, multiply Table 1 values of T_r by $T_t/290$.

(d) Opposite $\overline{NF}_{db}$ in first column, read T_e in fifth col.

T_a		T_a/L_r	
$1/L_r$		T_r	
$\overline{NF}_{db}$		T_e	
T_t		T_n	

(2) RANGE FACTORS		(3) DECIBEL VALUES	PLUS (+)	MINUS (-)
$P_{t(kw)}$		10 log $P_{t(kw)}$	.	.
$\tau_{\mu sec}$		10 log $\tau_{\mu sec}$	.	.
G_t		$G_{t(db)}$	.	.
G_r		$G_{r(db)}$	.	.
$\sigma_{50(sq.m.)}$		10 log σ_{50}	.	.
f_{Mc}		- 20 log f_{Mc}	.	.
T_n, °K		- 10 log T_n	.	.
$V_{o(50)}$		- $V_{o(50)(db)}$	.	.
C_B		- $C_{B(db)}$	.	.
L_t		- $L_{t(db)}$	.	.
L_r		- $L_{r(db)}$	.	.
L_p		- $L_{p(db)}$	.	.
L_x		- $L_{x(db)}$	.	.
Range-equation constant (40 log 1.292)			4.45	
4. Obtain column totals →			.	.
5. Enter smaller total below larger →			.	.
6. Subtract to obtain net decibels →			+ .	- .

7. In Table 2, find range ratio corresponding to this net decibel value, taking its sign (±) into account. Multiply this ratio by 100. This is R_o →

8. Multiply R_o by the pattern-propagation factor. F = $R_o \times F = R'$ →

9. On the appropriate curve of Figures 6 - 8, determine the atmospheric-absorption loss factor, $L_{a(db)}$, corresponding to R'. This is $L_{a(db)(1)}$ →

10. In Table 2, find the range-decrease factor corresponding to $L_{a(db)(1)}$, δ_1 →

11. Multiply R' by δ_1. This is a first approximation of the range, R_1 →

12. If R_1 differs appreciably from R', on the appropriate curve of Figures 6 - 8 find the new value of $L_{a(db)}$ corresponding to R_1. This is $L_{a(db)(2)}$ →

13. In Table 2, find the range-increase factor corresponding to the difference between $L_{a(db)(1)}$ and $L_{a(db)(2)}$. This is δ_2. →

14. Multiply R_1 by δ_2. This is the radar range in nautical miles, R_{50} →

Note: If the difference between $L_{a(db)(1)}$ and $L_{a(db)(2)}$ is less than 0.1 db, R_1 may be taken as the final range value, and steps 12 - 14 may be omitted. If $L_{a(db)(1)}$ is less than 0.1 db, R' may be taken as the final range value, and steps 9 - 14 may be omitted. (For radar frequencies up to 10,000 megacycles, correction of the atmospheric attenuation beyond the $L_{a(db)(2)}$ value would amount to less than 0.1 db.)

Acknowledgment

The information presented has been drawn from many sources, only some of which are referenced. Helpful and encouraging comment has been received from many persons, among whom the author especially thanks L. E. Davies of Stanford Research Institute, W. M. Hall of the Raytheon Company, and R. J. Adams of the Naval Research Laboratory.

References

[1] D. O. North, "An Analysis of the Factors which Determine Signal/Noise Discrimination in Pulsed Carrier Systems," RCA Labs., Princeton, N. J., Tech. Rept. PTR-6C; June 25, 1943.
[2] A. V. Haeff, "Minimum-detectable radar signal and its dependence upon parameters of radar systems," Proc. IRE, vol. 34, pp. 857–861; November, 1946.
[3] R. M. Ashby, V. Josephson, and S. Sydoriak, "Signal Threshold Studies," Naval Research Lab., Rept. R-3007; December 1, 1946. (This report is based on work done at the M.I.T. Radiation Lab., not all of which had been previously published.)
[4] K. A. Norton and A. C. Omberg, "The maximum range of a radar set," Proc. IRE, vol. 35, pp. 4–24; January, 1947.
[5] L. N. Ridenour, *et al.*, "Radar System Engineering," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1; 1947.
[6] R. Payne-Scott, "The visibility of small echoes on PPI displays," Proc. IRE, vol. 36, pp. 180–196; February, 1948.
[7] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 24; 1950. (Much of the information in this volume which is relevant to the present context is also contained in [3].)
[8] D. E. Kerr, *et al.*, "Propagation of Short Radio Waves," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 13; 1951.
[9] L. V. Blake, "The effective number of pulses per beamwidth for a scanning radar," Proc. IRE, vol. 41, pp. 770–774; June, 1953. "Addendum," Proc. IRE, vol. 41, p. 1785; December, 1953.
[10] W. M. Hall, "Prediction of pulse radar performance," Proc. IRE, vol. 44, pp. 224–231; February, 1956.
[11] H. D. Ko, "The distribution of cosmic background radiation," Proc. IRE, vol. 46, pp. 208–215; January, 1958.
[12] J. R. Bauer, *et al.*, "Radio Refraction in a Cool Exponential Atmosphere," M.I.T. Lincoln Lab., Lexington, Mass., Tech. Rept. No. 186, ASTIA Doc. No. 202331; August 27, 1958.
[13] B. R. Bean and G. D. Thayer, "Models of the atmospheric radio refractive index," Proc. IRE, vol. 47, pp. 740–755; May, 1959. P. Misme, B. R. Bean, and G. D. Thayer, "Models of the atmospheric radio refractive index," Proc. IRE, vol. 48, pp. 1498–1501; August, 1960. Also ,"CRPL Exponential Reference Atmosphere," NBS Monograph 4; October 29, 1959.
[14] L. V. Blake, "Radar Attenuation by Atmospheric Oxygen," paper presented at URSI-IRE Meetings, Washington, D. C., Commission 2, Tropospheric Radio Propagation; May 6, 1959. Also, "Curves of Atmospheric-Absorption Loss for Use in Radar Range Calculation," Naval Research Lab., Rept. 5601; January, 1961.
[15] R. C. Hansen, "Low noise antennas," *Microwave J.*, vol. 2, p. 19; June, 1959.
[16] J. J. Bussgang, *et al.*, "A unified analysis of range performance of CW, pulse, and pulse Doppler radar," Proc. IRE, vol. 47, pp. 1753–1762; October, 1959.
[17] D. C. Hogg and W. W. Mumford, "The effective noise temperature of the sky," *Microwave J.*, vol. 3, pp. 80–84; March, 1960.
[18] J. I. Marcum and P. Swerling, "Studies of target detection by pulsed radar," IRE Trans. on Information Theory, vol. IT-6; April, 1960.

Prediction of Coverage for Trans-Horizon HF Radar Systems*

G. F. ROSS†, senior member, ire, and L. SCHWARTZMAN†, member, ire

Summary—**Coverage patterns for trans-horizon radar systems can be predicted both analytically and empirically. Empirical methods are preferred because of their rapid application and lack of need for elaborate computing facilities. In addition, the coverage techniques introduced may be used eventually to determine the slant range, absorption, and effective noise temperature of trans-horizon radar systems. Some of the fundamental considerations which influence the prediction of coverage, such as geographical location, and the diurnal, seasonal, and anomalous behavior of the ionosphere, are discussed qualitatively. In spite of these factors, however, it is shown how HF radar systems overcome the line-of-sight limitations imposed upon conventional microwave radar systems and theoretically permit up to four or five times the range for a single hop.**

Introduction

THIS article presents some fundamental considerations influencing the prediction of the radar coverage for HF trans-horizon systems. For example, HF systems are dependent upon geographical position, and diurnal, seasonal and anomalous behavior of the ionosphere. These factors are described qualitatively with estimations of the anticipated "outage" time for the various regions of the globe.

Techniques are described which can be used to determine one-hop (over-the-horizon) skip distances in terms of the antenna beamwidth, launch angle, and "virtual" height for each of the respective rays in the beam. The interdependence of these parameters, along with the target range, are tabulated for convenient use by a trans-horizon radar system designer.

Several techniques for determining the virtual height as a function of launch angle, operating frequency, and assumed ionospheric model are presented. Two methods are analytical, employing parabolic and piece-wise linear approximations of the electron density profile, while a third technique employs empirical results obtained from ionospheric sounding stations.

The latter method makes use of the Smith transmission curves with sliding ionogram overlays, eliminating the long and detailed calculations. This method clearly demonstrates the cause of E and F_1 layer blanketing

* Received by the PGMIL, January 17, 1961.
† Surface Armanent Div., Sperry Gyroscope Co., Div. of Sperry Rand Corp., Great Neck, N. Y.

when F_2 layer propagation is desired. Radar coverages are obtained for typical winter days, winter nights, summer days, and summer nights.

The paper concludes with a qualitative discussion of the factors which influence transmission through the ionospheric region. Procedures are presented for graphically determining the normal "radar" spreading loss $(1/R)^4$ and the losses which result from propagation through an ionized region. It will be shown how the latter losses, together with the effective noise temperature, which are both functions of frequency, play an important role in the solution of the "radar equation."

The Ionosphere

To predict the coverage pattern for a trans-horizon radar, the designer should be thoroughly familiar with the many factors which influence and lead toward the determination of the coverage and the many ionsopheric anomalies that can play havoc with the predicted coverage pattern. However, before examining these anomalies and their effects, it is necessary first to be familiar with the problems associated with electromagnetic wave propagation in an ionized medium, *e.g.*, the ionosphere.

An isotropic and homogeneous region can be characterized by three constants, ϵ (the permittivity), μ (the permeability), and σ (the conductivity), which relate the fields in the following manner:[1]

$$\underline{D} = \epsilon \underline{E}; \qquad \underline{B} = \mu \underline{H}; \qquad \underline{J} = \sigma \underline{E}. \tag{1}$$

In such a medium, the phase and group velocities can be shown to be

$$v = \frac{c}{\sqrt{k_\epsilon k_m}}; \qquad k_\epsilon = \frac{\epsilon}{\epsilon_0}; \qquad k_m = \frac{\mu}{\mu_0}, \tag{2}$$

where c is the velocity of the wave in free space, and k_ϵ and k_m are the electric and magnetic specific inductive capacities, respectively. The ratio $\eta = c/v$ is called the index of refraction, and since k_m is usually nearly unity in all but ferromagnetic materials, the index of refraction is approximately equal to the square root of k_ϵ. The amount of bending that a wave undergoes in passing from one medium to another is described by Snell's law and is a function of the ratio of the indices of refraction of the two media, as well as the angle of incidence.

In an ionized medium, k_ϵ is frequency-dependent, which results in a functional relationship between the applied frequency and the phase velocity of the wave. Furthermore, the presence of a static magnetic field results in a birefringent medium. The effect is to split an incident wave into two components called the ordinary and the extra-ordinary rays. These rays follow different paths, having different phase velocities, and suffer different attenuations.

The ionosphere, to a first approximation, may be thought of as a sequence of ionized layers concentric with the earth and ranging in height from approximately 50 to 350 km. The layers are identified as D, E, F_1 and F_2 in order of increasing height and are distinguishable in terms of phenomena, many of which pertain to radio wave propagation. The heights of the layers are not fixed but depend upon the primary factors governing these phenomena, such as zenith angle of the sun, condition of the sun (with respect to sunspots, flares, etc.), and the earth's magnetic field and its changes. To some extent, the heights of the layers are functions of past as well as current events.

The ionic layers are present over the globe at all times, but there exist differences in appearance and behavior at different times and locations. Examples are diurnal and seasonal changes, variations between years of low solar activity and high solar activity, and between equatorial latitudes and polar latitudes. All of the phenomena change from time to time in a complicated manner not easily described except in statistical terms. The variations, both in magnitude and frequency, are least at the equator and greatest at the geomagnetic poles.

The reflection and refraction of waves in an ionized region depend upon the applied frequency, the direction of propagation with respect to the earth's magnetic field, the strength of the earth's magnetic field, and the electron density.

The complex index of refraction for such a medium is given by the Appleton-Hartree[2] equation as a function of the above parameters.

If one neglects the effects of collisions and the earth's magnetic field in the Appleton-Hartree expression and assumes that the change in electron density is small in the space of a wavelength, the index of refraction is given by

$$\eta = \sqrt{1 - \frac{8.1N}{f^2}}, \tag{3}$$

where N is the electron density in electrons per cubic centimeter/10^5 and f is the applied frequency in megacycles. Since the amount of bending or refraction experienced by a wave is related to n through Snell's law, a wave striking the ionosphere at vertical incidence will be returned to the earth when $\eta = 0$, or from (3) when

$$f_v = \sqrt{8.1N}. \tag{4}$$

The frequency given by (4) is called the vertical-incidence critical frequency and is used as a measure of electron density of the region. The height of a layer as well as the electron density is required to describe adequately the ionosphere. The height is usually given in terms of the "virtual height," which is defined as the

[1] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., ch. 5; 1941.

[2] S. K. Mitra, "The Upper Atmosphere," The Asiatic Society, Calcutta, India, 2nd ed.; 1952.

height of a reflecting layer in free space for which the travel time of the wave is equal to the travel time in the actual ionized medium.[3] The virtual height, as shown in Fig. 1, is always greater than the actual height, since the group velocity of propagation in an ionized region is less than the velocity of propagation in free space.

An important relationship concerning the virtual height is the equivalence theorem for a flat earth, a flat ionosphere, and without consideration of the earth's magnetic field. This relationship states that the virtual height of reflection of a wave propagated at oblique incidence over a path (the height of the equivalent triangular path) is equal to the virtual height of reflection for the wave of the equivalent vertical incidence frequency propagated vertically through the ionosphere. The equivalent vertical incidence frequency and the

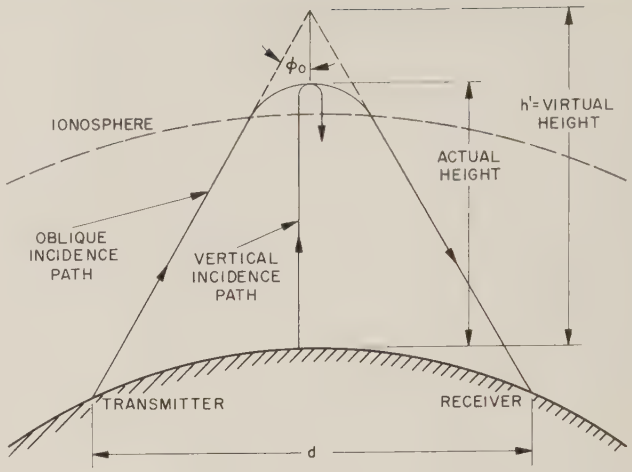


Fig. 1—Relation between virtual height and actual height.

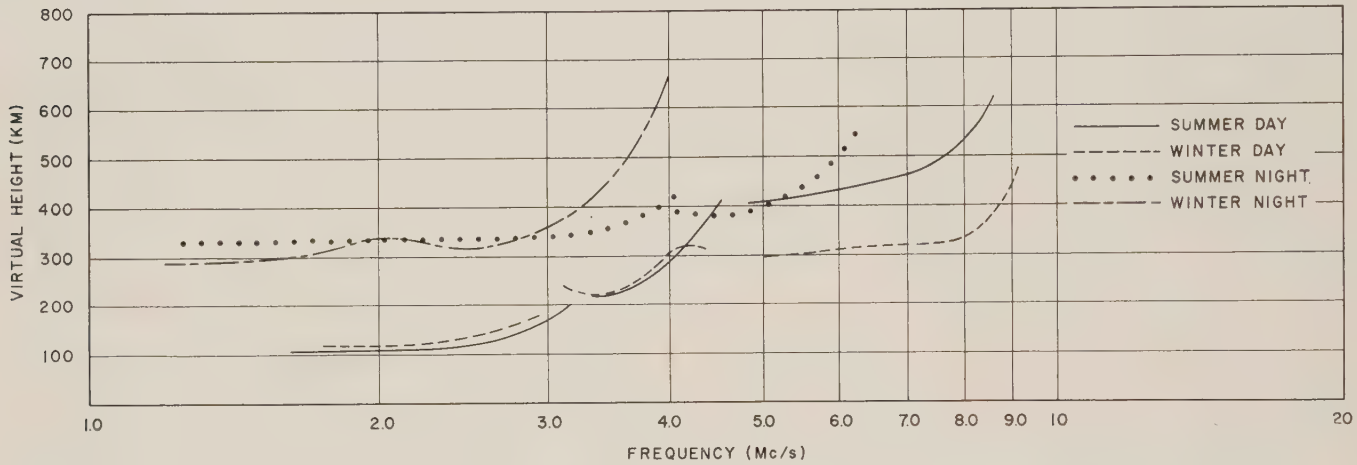


Fig. 2—Typical diurnal and seasonal ionograms.

oblique incidence wave frequency are related by the secant law, that is,

$$f = f_v K \sec \phi_0, \tag{5}$$

where f is the oblique-incidence frequency, f_v is the equivalent vertical-incidence frequency, ϕ_0 is the half angle at the apex of the equivalent path, and K is a correction factor assumed to depend only on the transmission path. The maximum oblique incidence frequency which will be supported (reflected back) by the ionosphere for a given distance of transmission is called the maximum usable frequency (MUF).

To determine the coverage pattern of an HF system, it is necessary to predict the path for the various rays in the beam. Since the ionosphere has a large anomalous behavior, *i.e.*, spread F, sporadic E, layer tilts, storms, etc., techniques which do not employ existing ionospheric data must be avoided since erroneous conclusions can be obtained. Incorrect results will also be obtained if data at only one point along the path are used.

For example, Fig. 2 shows four typical plots illustrating the large diurnal and seasonal variations of the critical frequencies and the virtual heights.[4] These plots are known as ionograms. In temperate latitudes, changes in the F_2 layer critical frequency (f_0F_2) are gradual, having a correlation time of approximately 20 to 30 minutes.[5] Wright, Zondt, and Stonehocher[5] have plots showing the variation of typical electron-density profiles as a function of latitude, longitude, and time of day.

In the low and middle latitudes, the ionosphere may be described as being either in a quiet or disturbed condition, but in the polar region, the latter condition usually prevails. Changes in f_0F_2 of as much as two to one within a 15-minute period are common during a winter day or night. In the low and middle latitudes, such changes are not usually encountered except during the onset of a storm.

There are two basically different types of storms

[3] F. D. Terman, "Radio Engineering," McGraw-Hill Book Co., Inc., New York, N. Y., 3d ed., p. 633; 1947.

[4] "Ionospheric Radio Propagation," U. S. Dept. of Commerce, Washington, D. C., NBS Circular No. 462, pp. 19–20; 1948.

[5] J. W. Wright, T. E. Van Zondt and G. H. Stonehocher, "Data on Ionospheric Electronic Densities," Boulder, Colo., NBS Rept. No. 5590; 1958.

which may profoundly alter the normal behavior of the ionosphere. They are the sudden ionospheric disturbance (SID) and the ionospheric storm. SID's occur only on the daylight side of the globe and are usually traceable to an event, such as a flare, occurring on the sun's surface. This disturbance may last from a few minutes to several hours and is characterized by a sudden severalfold increase in D-region electron density which produces above normal absorption of high-frequency signals. The lower frequencies are affected first, the higher frequencies last, with the higher frequencies being the first to recover. SID occurrence is related to the sunspot number[6] but usually is not predictable.

The ionospheric storm is of much longer lasting duration than the SID. An ionospheric storm is composed of two modes, the short time variation and the disturbance variation. The latter is dependent upon local time. It has been shown by many investigators that magnetic and ionospheric storms seem to be manifestations of the same fundamental phenomena. During magnetic storms, the maximum usable frequencies are lowered while the lowest usable frequencies rise, so that the operating bandwidth is reduced. Magnetic storms are most violent during and shortly after sunspot maximum. They have a tendency to occur at 27 day intervals (rotation of sun) and are most predictable during sunspot minimum years.[7] The occurrence of magnetic storms can be predicted six hours in advance with an 80 per cent probability, but the severity cannot be predetermined.

Finally, layer tilts, including sunrise and sunset effects have been known for a long time. Such tilts can deviate a wave off a great circle path and can cause bearing changes of 2° to 3° and elevation changes of about 5°.[8,9]

Ray Tracing

The general mathematical procedure for tracing the path of a wave through a nonisotropic medium is complicated and does not lend itself readily to application. Certain simplifying assumptions, however, yield approximate solutions which can be used by the radar designer to predict system coverage. These assumptions are:

1) The electron density and, hence, the index of refraction over narrow regions varies only with the radius r in spherical coordinates.
2) The equivalence theorem (as defined earlier) is valid for the case of a curved earth and curved ionosphere and the virtual height is the same as the height of the triangle formed by extending the straight-line portions of the wave paths, as shown in Fig. 1.
3) The wave can be represented by a series of rays drawn in the direction of the propagated energy and not affected by the earth's magnetic field. These rays are contained within the given antenna beamwidth.
4) Effects of refraction in the troposphere may be neglected.

With the acceptance of these assumptions, the general problem of radar coverage prediction is illustrated in Fig. 3 (next page). The height of the ionosphere is not shown to scale. Ray $\overline{ACD}$ was chosen to be tangent to the earth's surface at A and D, respectively, for expediency. The propagated energy is contained within the elevation beamwidth angle β and is refracted by the ionosphere in the region $\overline{BC}$. The rays are thus returned to earth in the region $\widehat{FD}$. It should be noted that rays launched at different angles penetrate differently into the medium. In the example, the tangent ray reaches a virtual height h_2' while the extreme ray extends to a height h_1'. The loci of the virtual bounce points of all rays lie on the path $\overline{BC}$.

A target is shown at H at a height Δ km above the earth's surface. By applying elementary principles of geometry and series approximation, the functional relationships between these distances and angles can be expressed and are shown in Table I. It is evident from these relationships that once the antenna beamwidth and virtual height are specified, the distances pertinent to the system design may be determined. These distances include measurements of

1) $\widehat{AJ}$, the ground range to a target located Δ km above the earth's surface,
2) $\widehat{FD}$, the illuminated ground region,
3) $\overline{BC}$, the corresponding illuminated region of the ionosphere,
4) $\widehat{AF}$ and $\widehat{AD}$, the range to the leading and trailing edge of the ground clutter or backscatter, respectively.

Determining the virtual height (h'), however, constitutes one of the major difficulties in ray tracing since this distance is a complicated function of the wave angle of entry into the medium, the electron density, and the transmitted frequency. Several procedures, both analytical and empirical, have been devised to find h'. Analytical methods, for example, involve choosing a model ionosphere with either a parabolic or piece-wise-linear electron density distribution and applying Snell's law at the boundaries.[10–12] An example of this procedure

[6] A. H. Shapely and J. V. Lincoln, "Some Methods for General Prediction of S.I.D.," U. S. Dept. of Commerce, Washington, D. C., NBS, Rept.; 1947.
[7] NBS Circular No. 462, *op. cit.*, pp. 63–66.
[8] H. A. Whales, "Effective tilts of the ionosphere at places about 1000 km apart," *Proc. Phys. Soc.*, vol. 69, pt. B., pp.301–310; January-December, 1956.
[9] S. Stein, "The role of layer tilts in long range HF radio propagation," *J. Geophys. Res.*, vol. 63, pp. 217–241; March, 1958.
[10] A. H. DeVoogt, "The calculation of the path of a radio ray in the ionosphere," Proc. IRE, vol. 41, pp. 1183–1186; September, 1953.
[11] A. H. DeVoogt, "Ionosphere models as an aid for calculation of ionospheric propagation quantities," Proc. IRE, vol. 48, pp. 341–345; March, 1960.
[12] D. B. Muldren, "An ionospheric ray tracking technique and its application to a problem in long distance radio propagation," IRE Trans. on Antennas and Propagation, vol. AP-7, pp. 393–396; October, 1959.

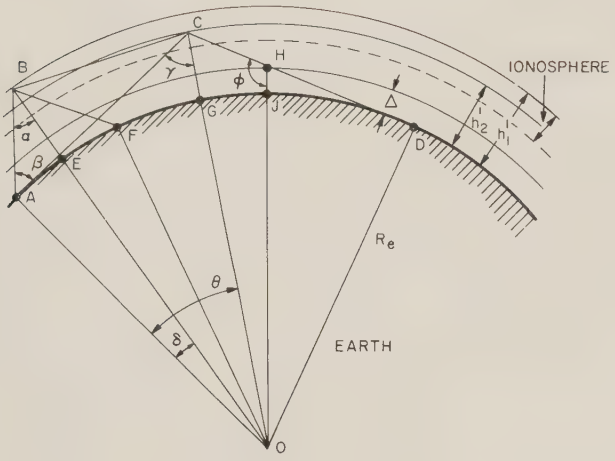


Fig. 3—HF radar geometry.

TABLE I
TABLE OF FORMULAS

Distance (statute miles)	Relationships (all angles in degrees)
$\widehat{AJ}$	$69(2\theta-\phi^1)$, $\phi^1=\phi-90$
$\widehat{FD}$	$138(\theta-\delta)$
$\overline{BC}$	$7926(1+x+\bar{x})^{1/2}\left(\sin\frac{[\theta-\delta]}{2}\right)$
$\widehat{AF}$	138δ
$\widehat{AD}$	138θ

Where

$\theta=\cos^{-1}(1-\bar{x}+\bar{x}^2\cdots)$, $\bar{x}=h_2{}^1/6377$, $h_2{}^1$ in km
$\alpha=\sin^{-1}[(\cos\beta)(1-x+x^2\cdots)]$, $x=h_1{}^1/6377$, $h_1{}^1$ in km
β, launch angle for extreme ray
$\delta=90-\alpha-\beta$
$\phi^1=\cos^{-1}[(\cos\theta)(1-y+\bar{x})]$, $y=\Delta/6377$, Δ in km

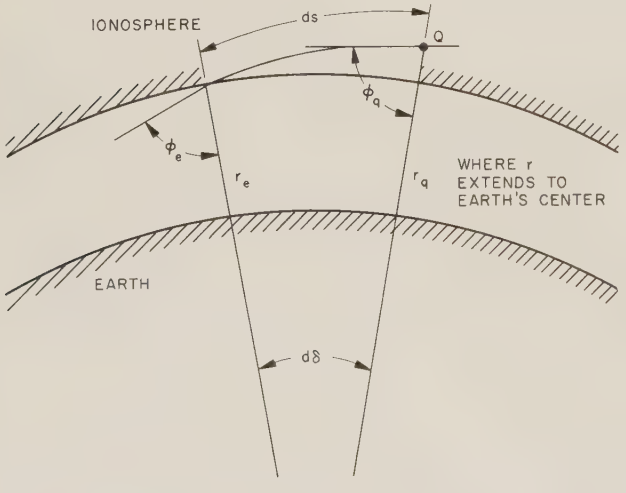


Fig. 4—Snell's law at a boundary.

may be illustrated with the aid of Fig. 4. At the troposphere-ionosphere boundary,

$$r_q\eta_q \sin\phi_q = r_e \sin\phi_e, \tag{6}$$

where η is the index of refraction of the medium.

The index of refraction of the ionosphere is a function of transmitted frequency and electron density and is given by (3). The electron density (N) is, of course, a function of the height above the earth's surface.

If N is assumed described by the following relationship [shown in Fig. 5(a)],[10]

$$kN = a + b/r + g/r^2, \tag{7}$$

where a, b, and g are constants to be obtained from the boundary conditions, then

$$\eta^2 = 1 - kN = 1 - (a + b/r + g/r^2). \tag{8}$$

For $r=r_e$, N is zero. This is equivalent to stating that the ionosphere begins at a discrete altitude. That is,

$$ar_e^2 + br_e + g = 0. \tag{9}$$

Assuming the ionization starts smoothly at the boundary,

$$dN/dr = 0. \tag{10}$$

If (7) is differentiated and set equal to zero, and if the result is substituted into (9), the following is obtained:

$$r_e = -2g/b \quad \text{and} \quad 4ag = b^2,$$

or

$$b = -2ar_e \quad \text{and} \quad g = ar_e^2. \tag{11}$$

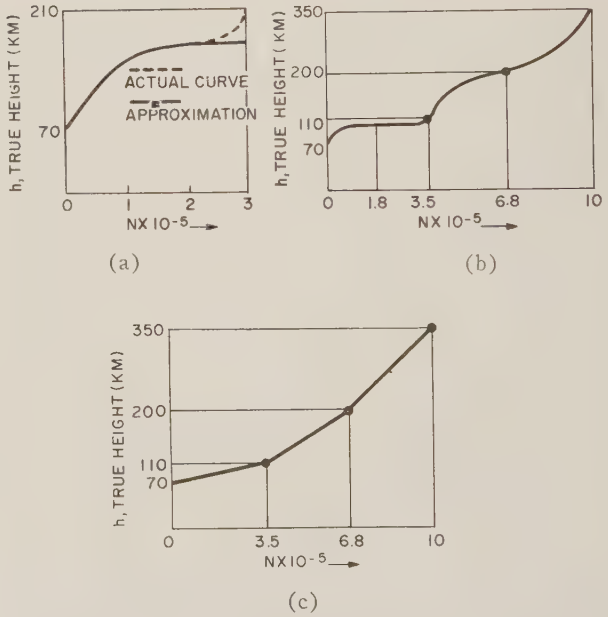


Fig. 5—Electron density profiles. (a) Parabolic approximation. (b) Confluent parabolic approximation. (c) Piece-wise linear approximation.

If (11) is substituted into (7) and solved for N,

$$N = \frac{a}{k}\left(1 - \frac{r_e}{r}\right)^2. \tag{12}$$

To find the apex or turning point of the ray, the angle ϕ_q in Fig. 4 is set equal to 90° and substituted into (3) and (6):

$$\eta^2 = \left(\frac{r_e \sin \phi_e}{r_{\max}}\right)^2 = 1 - kN = 1 - a\left(1 - \frac{r_e}{r_{\max}}\right)^2. \tag{13}$$

If the expression is solved for $r_{\max}$ at the turning point,

$$r_{\max} = r_e\left(\frac{a + \sqrt{\sin^2 \phi_e + a \cos^2 \phi_e}}{a - 1}\right). \tag{14}$$

Using the value of $r_{\max}$ given by (14) and the technique outlined by DeVoogt[10] to obtain the central angle δ (see Fig. 3), the virtual height is easily found. Note that the angle ϕ_e is geometrically related to the launch angle β, while the constant a and the radius r_e depend upon the selected ionospheric model and operating frequency. The height attained by other rays in the beam can be determined in similar fashion since the only variable in (14) is the angle ϕ_e.

A more accurate analysis might be obtained by approximating the electron density-altitude profile by a series of confluent parabolas as shown in Fig. 5(b), where the end-point conditions at each boundary must be continuous in magnitude and slope. The calculations for more than two boundaries become cumbersome and require the use of computer facilities.

The second analytical approach for approximating the N-h' profile employs the use of piece-wise-linear segments in lieu of the confluent parabolas discussed above, and is shown in Fig. 5(c). Although this technique initially introduces simplification into the equations, a large number of line segments are required to achieve a desired accuracy. This requires the matching of a great many boundary conditions. The piece-wise-linear approach is, however, particularly adaptable to programming in a digital computer.

It should be clear from the preceding discussion that analytical methods which lead to the determination of virtual height (or actual path length) are quite involved and not very tractable. The next technique to be described is empirical in nature and requires the use of Smith transmission curves with sliding ionogram overlays (Fig. 6).[13] The Smith curves have the advantage of yielding rapid and reasonable accurate coverage information and appear to be the most generally accepted method for field application.

The transmission curves are made on a transparency with the virtual height scale identical to that of the ionogram. The frequency axis of the ionogram is the

[13] NBS Circular No. 462, *op. cit.*, pp. 69-74.

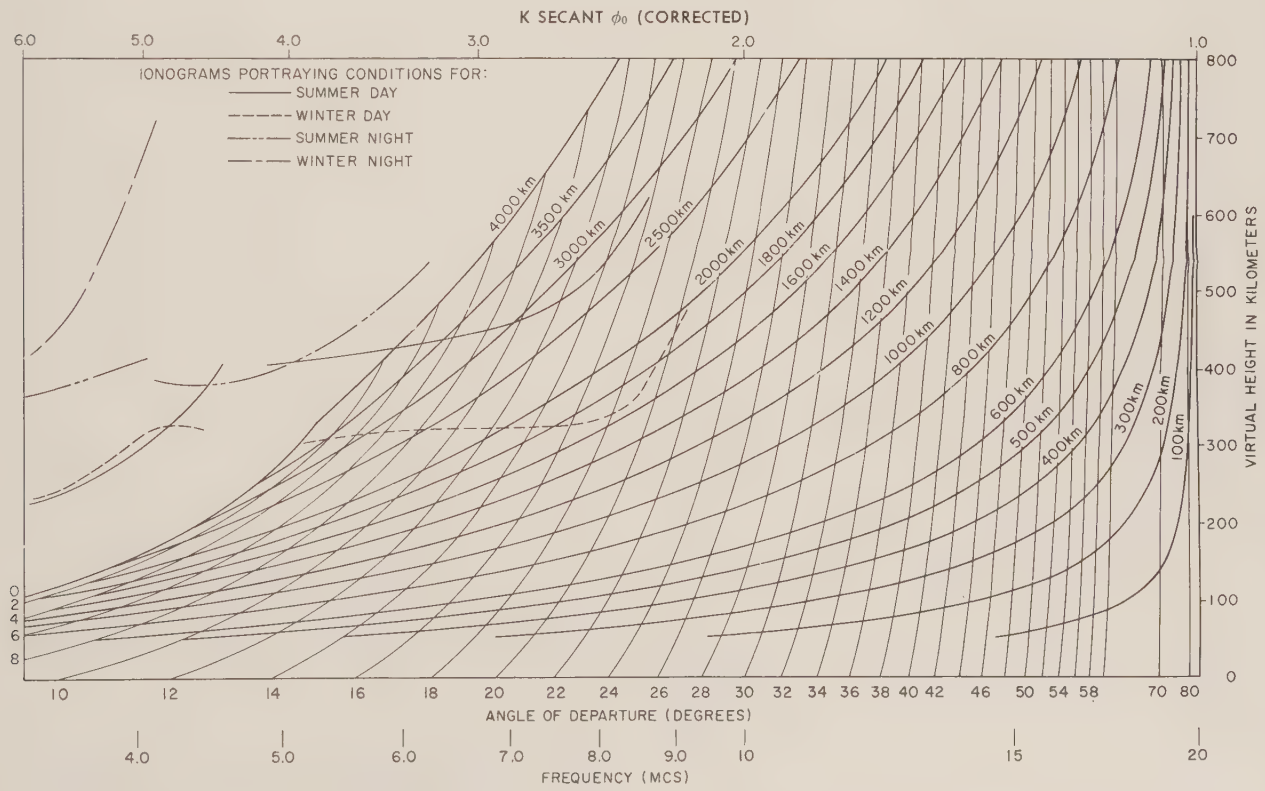


Fig. 6—Overlay of Smith transmission curves on ionograms.

same logarithmic scale as the K sec ϕ_0 (*i.e.*, sec ϕ_0 corrected) scale of the transmission chart, but is plotted in the opposite direction.[14] These two scales are then identical to the CI and D scales of a slide rule and allow the multiplication of the vertical incidence critical frequency with K sec ϕ_0 giving the operating frequency [see Fig. 6]. The number 1.0 on the K sec ϕ_0 scale is the index used for this process. If the operating frequency is chosen and one wishes to find the coverage, the 1.0 index on the transmission chart is aligned with the operating frequency. In this manner, the radiation angle and distance intersection with the various layers are quickly and easily obtained. For only F_2 layer propagation, the angles of departure can be read from the abscissa of the Smith curves, and the ground coverage determined from the parametric distance curves. For example, assume that the operating frequency is 20 Mc and only the F_2 layer is to be used. In Fig. 6, the ionograms from Fig. 2 are overlayed on the transmission chart with the 1.0 index placed on 20 Mc. It is then seen that the intersection of the angle-of-departure curves with the various layers occurs only for a winter and summer day. That is, the frequency is too high for either a winter or summer night and the rays will therefore "skip." For angles of departure between 4° and 18° on a winter day, the ground coverage will be from 1600 km to approximately 3100 km.

E-layer shielding of the F_2 layer is also quickly determined by noting the radiation angle curve. If the same radiation angle passes through both the E and F_2 layer, then the E layer shields the F_2 layer at this angle.[15]

Prediction of Coverage

To illustrate the method of predicting radar coverage, use will be made of the ionograms and the Smith transmission curves shown in Fig. 6. These rather general ionospheric conditions were chosen to examine the variations in trans-horizon radar system coverage throughout the year. The ionogram information from Fig. 2 has been simplified to exclude the data obtained from multiple reflections and that due to the earth's magnetic field, *i.e.*, the extraordinary wave. By sliding the Smith transmission chart over these ionograms in the manner described in the preceding paragraphs, it is possible to determine ground-range coverage for F_2 layer propagation as a function of frequency and launch angle for the seasonal and diurnal conditions shown in Fig. 2. This information is plotted in Figs. 7 and 8 and may be interpreted with the aid of Fig. 3 in the manner described below.

For an operating frequency of 14 Mc on a summer day, the illuminated ground region $\widehat{FD}$ is approximately 900 km (2500 minus 1600 km). The effective

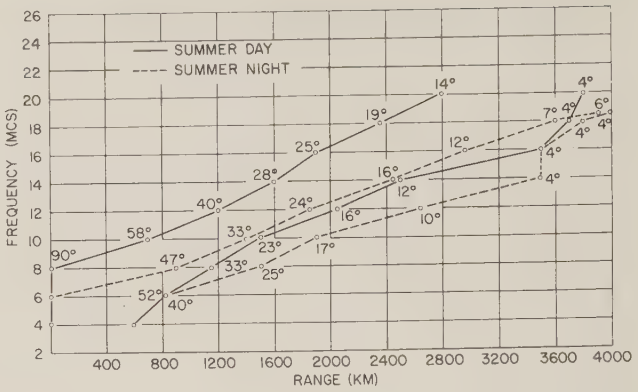


Fig. 7—Range coverage for summer day and summer night.

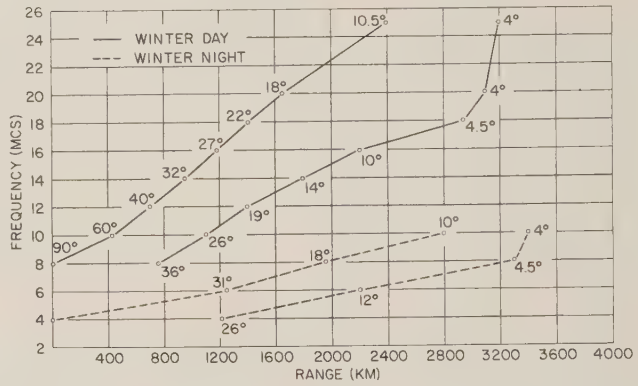


Fig. 8—Range sector coverage for winter day and winter night.

beamwidth β' is 16° (28° minus 12°). Rays launched at angles exceeding β' will "skip" through the ionosphere and will not return to earth, while rays below 12° will be refracted by the E or F_1 layers and will be returned somewhere in the $\widehat{AF}$ region.

It can also be noted from these curves that single-hop system range is limited by ionospheric conditions on a winter day to 3200 km. The narrow and changing effective beamwidth β', resulting from the changing ionospheric conditions, necessitates a frequency sweep capability to cover the complete range $\widehat{AJ}$. Figs. 7 and 8 indicate that the transmitted frequency might be typically in the band from 4 to 25 Mc.

Application to HF Radar Systems

The slant range to targets can be found by using the equivalent triangle approximation established earlier and illustrated in Fig. 9. The exact computation of path length has been shown by Terman[16] to be unwieldy, and usually requiring the solution of an elliptical integral for even the simplest electron density distributions.[17]

[14] *Ibid.*, p. 72.
[15] *Ibid.*, p. 74.

[16] F. D. Terman, "Radio Engineers Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., 1st ed., p. 716; 1943.
[17] W. G. Baker and C. W. Rice "Refraction of short waves in the upper atmosphere," *Trans AIEE*, vol. 45, p. 302; February, 1926.

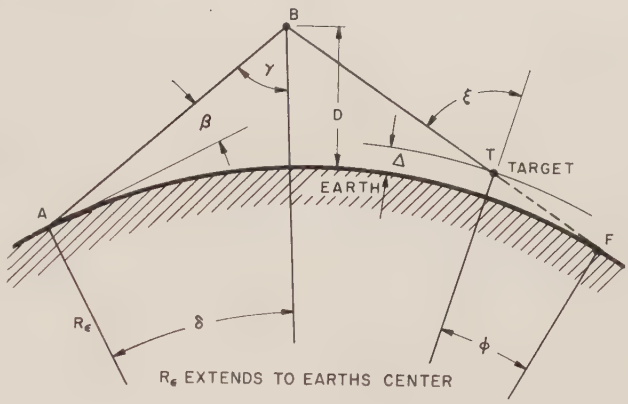
The slant range determined by the equivalent triangle approximation is used to calculate the inverse fourth power or radar-spreading loss. Calculating this loss brings into evidence one of the interesting differences between conventional microwave and HF radar systems. In HF systems, the propagation path length or slant range used in the radar equation changes as a complicated function of the launch angle, the operating frequency, and the conditions of the ionosphere. That is, the slant range to a target whose location with respect to the radar is fixed varies continually with changes in the ionospheric conditions.

The angle of ray entry into the ionized region (ϕ_e in Fig. 4) is also of importance to the system designer, since ionospheric abosorption or attenuation takes place mainly in the D layer and increases with increasing angle ϕ_e for a fixed frequency. Absorption generally depends upon the 1) year, season, and time of day, 2) operating frequency and ray entry angle, 3) geographical location, and 4) ionospheric conditions.

These factors determine the sun spot number, solar zenith angle, absorption index, and gyromagnetic frequency from which the total absorption may be calculated using the Appleton equation.[18] Absorption losses decrease essentially to zero during evening hours while attaining a high during summer afternoons; absorption also decreases as an inverse function of transmitted frequency. In addition to absorption losses, a polarization loss of approximately 8 to 9 db is usually assumed for ionospheric propagation[18] in the HF band.

Another interesting characteristic of HF radar systems is the effect of the frequency sweep on the required transmitted power. Earlier it was shown that a frequency-swept signal was necessary to obtain the extensive range coverage of an HF system. It was also noted that slant range to a target did not decrease linearly with decreasing ground range. As the frequency is decreased, however, the effective noise temperature increases so rapidly that it may actually compensate for the reduction in the inverse fourth power loss over a portion of the coverage region. Hence, the power requirements for detection of a target with a given cross-sectional area only decrease slowly as a function of decreasing target range.

Finally, there is the problem of separating the target echo from the backscatter or clutter return. In HF systems this problem is more acute than that encountered in the conventional microwave radar. Physical antenna size limitations, for example, restrict the minimum antenna beamwidths, and thus, large areas of the earth's surface are illuminated as illustrated in Fig. 10. By sweeping in frequency, it is possible to place the target in the leading ray of the beam ($\overline{LFC}$) and therefore outside the clutter region. This, however, restricts the maximum range of the system. Specifically, Fig. 10 depicts the conditions for a winter day with the system frequency



$$\overline{ABF} = 2R_\epsilon \frac{\sin\delta}{\sin\gamma},$$

where

$$\gamma = \sin^{-1}\left[\cos\beta\left(1 - \frac{D}{R_\epsilon} + \left(\frac{D}{R_\epsilon}\right)^2 \cdots\right)\right],$$

and $\delta = 90° - \beta - \gamma$.

$$\overline{ABT} = \overline{ABF} - \overline{TF} = 2R_\epsilon\left[\frac{\sin\delta}{\sin\gamma} - \frac{1}{2}\frac{\sin\phi}{\sin\xi}\right],$$

where

$$\xi = \sin^{-1}\left[\cos\beta\left(1 - \frac{\Delta}{R_\epsilon} + \left(\frac{\Delta}{R_\epsilon}\right)^2 \cdots\right)\right]$$

and $\phi = 90° - \beta - \xi$.

Fig. 9—Determination of slant range.

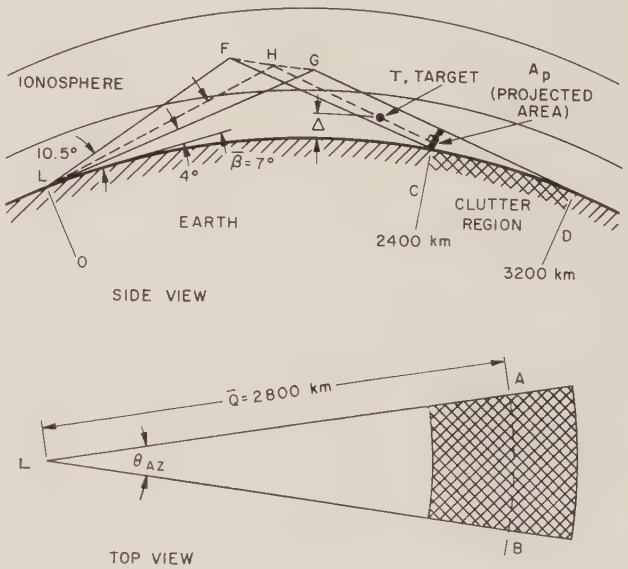


Fig. 10—Determination of clutter.

[18] P. Laitmen and G. Hayden, "Analysis and Prediction of Sky Wave Field Intensity in the HF Band," U. S. Army Signal Corps Radio Propagation Agency, Ft. Monmouth, N. J., Rept. No. RPU203; March, 1956.

chosen for maximum range. The projected area is given approximately by $A_p = \widehat{AB} \cdot \widehat{CD} \sin \beta$. The equivalent radar cross section of the clutter (σ) depends upon the characteristics of the illuminated surface and the transmitted frequency and is related to A_p by a coefficient K(*i.e.*, $\sigma = KA_p$). For example, water, which exhibits almost specular reflection characteristics, has a coefficient K equal to 10^{-4} or 10^{-5}, while for a poor earth this factor might approach unity.[19,20] Numerically, the clutter cross section, even over water, is enormous, and for the case shown in Fig. 10, is approximately 10^6 square meters. Equivalent target and clutter cross sections, however, should be compared only at the same ranges, so that besides the coefficient K, another factor should be included to account for the fourth-power loss over approximately the distance $\overline{TC}$.

Conclusions

Analytical methods to predict the radar coverage for HF trans-horizon systems have been described. These methods utilize ionospheric models which must be continually modified to account for the many anomalies. The calculations for each model are lengthy and require elaborate computer facilities. The empirical technique using the Smith transmission curves with ionogram overlays is to be desired, since it eliminates the need for detailed calculations, while providing the coverage predictions on a real time basis.

Some of the principal differences between the conventional microwave and HF radar systems were discussed. It was shown that the predicted coverage for HF radar systems is dependent upon geographical position, and the diurnal, seasonal, and anomalous behavior of the ionosphere. The inability to forecast the ionospheric anomalies such as spread F, sporadic E, SID, tilts, etc., can alter the mode of propagation and cause large errors in the predicted coverage. On the other hand, HF systems overcome the line-of-sight limitations imposed upon conventional microwave systems and theoretically permit up to four or five times the range for a single hop.

Acknowledgment

The authors express gratitude to the U. S. Air Force Rome Air Development Center, without whose support this paper could not have been written. Gratitude is also extended to A. Abston of the Sperry Gyroscope Company for his able assistance in the preparation of the paper.

[19] E. D. R. Shearman, "The technique of ionospheric investigation using ground backscatter," *Proc. IEE*, vol. 103B, p. 210; October, 1955.
[20] "A Summary of Literature Pertaining to HF Backscatter," Staff Commun. and Propagation Lab., Stanford Res. Inst., Palo Alto, Calif., NONR 225(33); 1960.

A New Display for FM/CW Radars*

HERBERT H. NAIDICH†

Summary—**One of the limitations of FM/CW radars, often quoted in the literature, is in their inability to handle multiple targets. Several variations of a comparatively simple Range/Range Rate Display, capable of handling and resolving a large number of targets when used in conjunction with an FM/CW radar, are presented. The FM/CW plane and means for resolving target ambiguities are discussed, as well as system features, which, when used in conjunction with the displays, enable multiple target range and velocity information to be obtained and utilized simultaneously.**

I. Introduction

THERE has been a great deal of interest in recent years in improving the operating characteristics of radars by means of techniques more sophisticated than those of pulse echoing. To achieve maximum range, a radar system must have a high average transmitted power and a minimum noise figure in the receiver. Maximum target resolution and range accuracy are proportional to the bandwidth of the modulation of the carrier and the ability of the receiver to process the intelligence contained within this modulation. Lastly, the longer the "time on target" of the radar energy, the more exactly the target velocity can be determined from the Doppler information contained in the received signal.

In pulsed radars, increasing accuracy and resolution by reducing pulsewidth (and increasing harmonic content of the modulation) causes problems in obtaining maximum average power on target and in the utilization of "time on target" to obtain the most accurate velocity determination. The manipulation of the pulse width, average power, and "time on target" factors

* Received by the PGMIL, January 13, 1961.
† IT & T Federal Labs., Nutley, N. J.

for pulsed systems results in compromises which often degrade some of the operating characteristics of the radar. FM/CW radars can often be designed with less compromise to performance specifications, but they encounter complex data-processing problems, especially if multiple targets are considered.

One advantage of pulsed radar systems is the range-echo-time relationship, which enables great simplicities in the data processing of the returned signals. Many types of CRT displays have been used based on the deflection of an electron beam in synchronism with the transmitted energy. When the echo is received, the electron beam is intensified forming a "blip" (PPI, RHI), or deflected forming a "pip" (Type A, J). The echo time (12.2 microseconds per nautical mile) is thereby reduced to a displacement on the face of the CRT and presented to an operator in a manner most applicable to the function of the particular radar. This simple echo-time relationship is one of the major reasons for the extensive utilization of pulsed radar systems. In FM/CW radars, where this echo-time relationship cannot be used, the processing of the information contained in the recovered modulation has been quite complex. Because of this difficulty, the use of FM/CW radars generally has been restricted to systems where multiple targets are not considered, thus simplifying the data-processing requirements. The radio altimeter is one type of single-target FM/CW equipment which has been in use for many years. The ground echo in this case is the only target under consideration.

It is the object of this paper to present a new, recently developed, display, which when used with an FM/CW radar system can process the intelligence contained in the received echo in a simple manner and present it to an observer in the form of a range/velocity display.

II. The FM/CW Equations

Linear frequency modulation of the radiated energy will be considered although other types have been used in the past. Fig. 1 illustrates various fundamental types of modulation that can be used. These types can be combined to obtain more complex forms of modulation. Fig. 1(d) is often used and is a combination of the modulations in Figs. 1(b) and 1(c). The basic equations from which it is possible to determine the target's velocity and position for the various types of modulation are:

Type A (unmodulated carrier):

$$F_A = f_d = K_1\dot{r}, \tag{1}$$

Type B (negative frequency modulation):

$$F_B = f_d + f_r = K_1\dot{r} + K_2 r, \tag{2}$$

Type C (positive frequency modulation):

$$F_C = f_d - f_r = K_1\dot{r} - K_2 r, \tag{3}$$

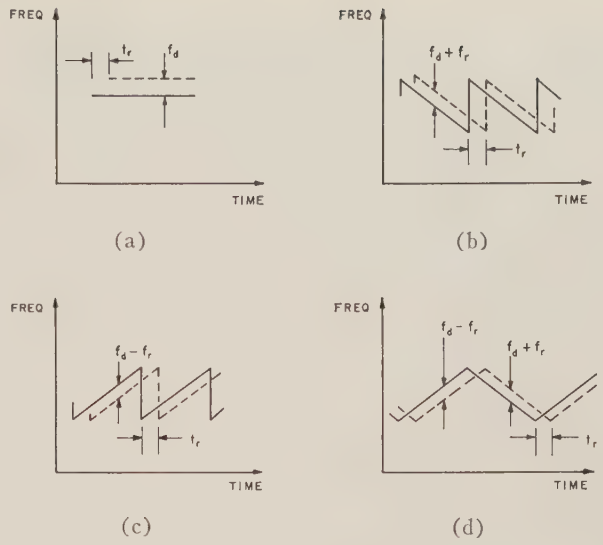


Fig. 1—Various frequency-modulation methods.

where F_a is the frequency displacement of the returned echo from the carrier when the transmitted wave is unmodulated [Fig. 1(a)]. This displacement is, of course, due to the Doppler effect $f_d = K_1\dot{r} = (2F_0/C)\dot{r}$. K_1 is the velocity-sensitivity factor and is established by the carrier frequency F_0 and the velocity of propagation C. $\dot{r}$ is the radial velocity of the target and is considered positive for inbound targets.

F_b is the difference in frequency between the transmitted and received frequencies when the transmitted wave is linearly frequency modulated in a negative direction and the measurement is made at the time of reception [see Fig. 1(b)]. This frequency difference has two components, f_d and f_r. f_r is the amount of frequency shift that would be obtained if the target were stationary during "B" type modulation and is due to the delay time t_r of the echo. In this type of modulation the echo would always be higher in frequency than the transmitted wave. f_r is proportional to the range by the range-sensitivity factor $K_2 = 2\dot{F}/C$, where $\dot{F}$ is the rate of frequency change of the carrier, and C is the velocity of propagation. If the target is moving, the Doppler shift f_d would cause an additional shift in received frequency.

During type "C" modulation, positive frequency deviation of the carrier, the received frequency for a fixed target would always be below the carrier, hence the negative sign in (3). The Doppler shift f_d would cause a higher frequency return for inbound targets. Fig. 1(c) illustrates this type of modulation. The range-sensitivity factor is identical if the rate of frequency change were the same as in type "B" modulation.

For a single target, only two of the three equations are required to obtain the necessary data to solve for the range and radial velocity of the target, and the triangular modulation of Fig. 1(d), combining Figs. 1(b)

and 1(c) would be suitable. This type of modulation is not adequate for multiple-target operation because of the resultant ambiguities due to the fact that the target returns during the "B" modulation—F_{b1}, F_{b2}, F_{b3}, $\cdots$, F_{bn}, and the returns during "C" modulation—F_{c1}, F_{c2}, F_{c3}, $\cdots$, F_{cn} are not correlated. For n target there will be n^2 possible solutions for the targets' range and velocity coordinates. If a third redundant mode is used, the N^2-N target ambiguities can be eliminated. For simplicity, the third mode will be considered to be type "A" unmodulated operation. Type "B" or "C" modulation with different range-sensitivity factors K_2 could also be used with possible advantage.

III. The FM/CW Plane

If a plane is constructed where the ordinate is the range or f_r axis and the abscissa is the velocity or f_d axis, the resultant plane can be considered to be the FM/CW plane (Fig. 2). A target's coordinates f_d, f_r would be plotted in the first quadrant if inbound, and in the fourth quadrant if outbound. In Figs. 2(a)–(d), two targets are located in the FM/CW plane with the coordinates f_{d1}, f_{r1} and f_{d2}, f_{r2}.

Fig. 2(a) represents the locus of (1). For unmodulated or "A" type modulation, the target information F_{a1}, F_{a2} plots as two vertical lines with f_d axis intercepts at F_{a1} and F_{a2}. Fig. 2(b) depicts the locus of the two targets when type "B" modulation is used. The equation $F_b=f_d+f_r$ indicates that the locus of the two targets is on two lines with a slope of -1 with f_d and f_r axis intercepts of F_{b1} and F_{b2}, respectively.

Similarly, for "C" type modulation the locus of the targets must be on the two lines with a slope of $+1$ and f_d and f_r axis intercepts of F_{C1} and F_{C2} [see Fig. 2(c)]. Fig. 2(d) illustrates the superposition of these lines. The points of triple intersection are the coordinates of the two targets f_{d1}, f_{r1} and f_{d2}, f_{r2}. Thus, it is seen that a graphical solution to the three basic FM/CW equations can be achieved. Plotting the three sets of returns as indicated using, F_{a1}, F_{a2}, F_{a3}, $\cdots$, F_{an}; F_{b1}, F_{b2}, F_{b3}, $\cdots$, F_{bn}; and F_{c1}, F_{c2}, F_{c3}, $\cdots$, F_{cN}, results in N triple intersections at the proper target coordinates. From this graph, it is clear that removal of one of the modes—the A mode in this case—would result in four solutions, or n^2-n ambiguities. A solution by means of a CRT display seems to be indicated.

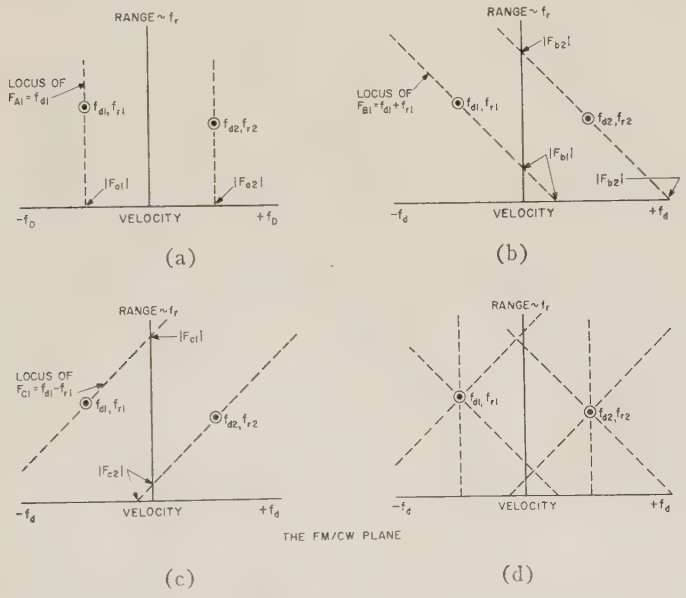


Fig. 2—Graphical solution of the range-velocity coordinates of multiple targets.

IV. Panoramic Display

While other types of spectrum analyzers are available, let us assume that three filter banks are used for frequency determination, one for each mode. The output of each filter is connected to a detector and storage element. The A, B, and C filter banks are sequentially scanned, and in synchronism with the filter scan, a spot is deflected from left to right on a CRT. If the start of the B and C filter scans are successively displaced vertically, as indicated in Fig. 3, and a vertical deflection is induced when a filter is reached whose storage element is energized, a single target will appear as indicated in Fig. 3. More than one target could be present, but only one will be considered here. The resultant spectrum display is quite conventional and can be achieved by a number of other well-known methods.

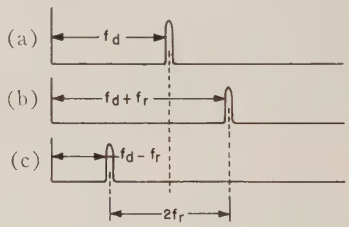


Fig. 3—Frequency spectrum for single target.

V. The FM/CW Display

The graphical solution of Fig. 2 can be electronically implemented by means of the FM/CW plane display, using the same signals required for the panoramic display. The following is the sequence of events for operation of the FM/CW raster. During the A mode, the horizontal deflection is from left to right, in synchronism with the A-filter readout. Simultaneously, the vertical deflection plates of the CRT are driven by a sawtooth whose period is equal to the sampling time per filter. By this means, a series of vertical lines is generated, one for each filter. During the B mode, an equal number of lines with a slope of -1 is generated. This is accomplished by using the identical wave shapes used for deflection during the A mode, but adding to the hori-

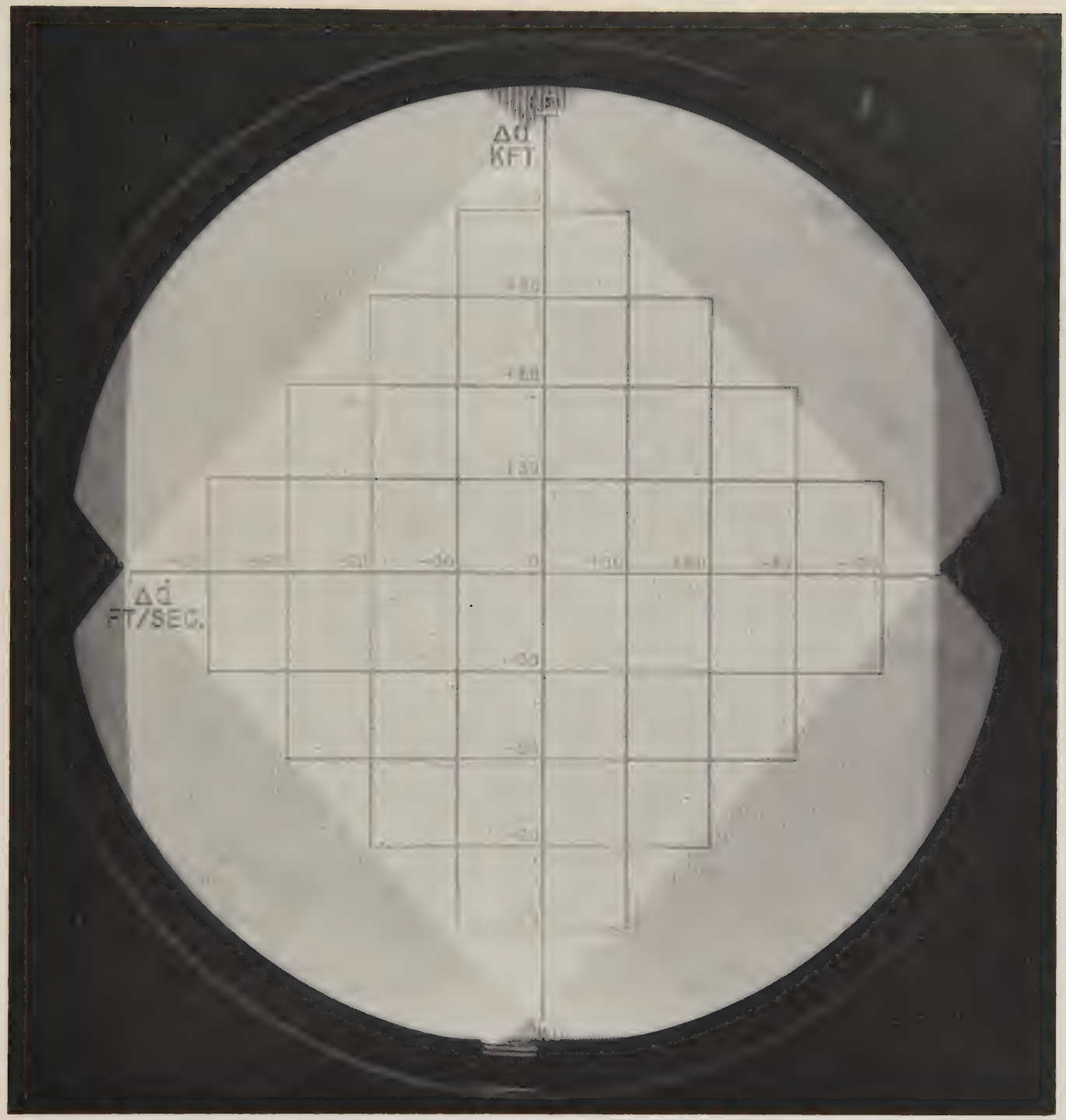


Fig. 4—FM/CW display raster.

zontal deflecting sawtooth wave shape a sawtooth which is identical to the one which is used for vertical deflection. Reversing the polarity of this added sawtooth during the *C*-mode scan causes an identical number of lines with a slope of +1 to be generated. With suitable persistence in the CRT phosphor, an FM/CW display raster is generated. Fig. 4 is a photograph of such a raster generated on a twelve-inch electrostatically-deflected CRT. If the lines are intensified only when an energized filter is reached, the resultant display will contain triple intersections whose coordinates are proportional to range and velocity of the target. The signal used for vertical "pip" deflection in Fig. 3 is used for intensification of the FM/CW plane display.

VI. The FM/CW Sector Display

It is quite often impractical to supply enough filters to cover the entire range of anticipated target fre-

quencies. This is especially true if the filter bandwidths are made extremely narrow in order to achieve high resolution. Filter banks covering only a limited spectrum can be used if variable-frequency oscillators are used to heterodyne the portion of the full spectrum that is of interest to the frequencies covered by the particular filter bank. Fig 5 illustrates the coverage on the FM /CW plane of the A, B and C filter banks when a limited number of filters are used and the heterodyning frequencies f_{ha}, f_{hb}, and f_{hc} are properly interrelated. The diamond-shaped area of mutual coverage is the only place on the FM/CW plane where it is possible to obtain a triple intersection. If the heterodyning frequencies are not properly interrelated, the coverage will be reduced. For instance, if only the heterodyning frequency used during the A mode f_{ha} is changed in frequency, the vertical lines, which indicate A-bank coverage, will move horizontally reducing the mutual coverage. If this frequency change is sufficiently large there will be no mutual coverage. When properly interrelated the diamond-shaped area can be considered to be a "window" on the entire FM/CW space, and this window can be positioned to cover the coordinates of the area on the FM/CW plane where the target is expected to appear. Several methods are available for achieving the correct relationship between the three heterodyning frequencies.

If it is desired to position the A, B and C filter banks so that a particular target with the coordinates f_d, f_r appears in the center of the three filter banks, the three heterodyning frequencies required are

$$f_{ha} = f_{IF} + f_f + f_d \tag{4}$$

$$f_{hb} = f_{IF} + f_f + f_d + f_r \tag{5}$$

$$f_{hc} = f_{IF} + f_f + f_d - f_r, \tag{6}$$

where f_{IF} is the carrier frequency of the IF before the final mixer, and f_f is the center frequency of the filter banks.

Fig. 6 is a block diagram of a system for obtaining these frequencies. By means of range and velocity controls, the center of the FM/CW sector coverage can be positioned to any place on the FM/CW plane. The range VFO moves the diamond-shaped area vertically and the velocity VFO positions this area horizontally. By measuring these two frequencies and using suitable scale factors, the center coordinates in range and velocity of the area of coverage can be displayed on a digital readout device. The FM/CW plane display is synchronized exactly as in the case of the full FM/CW display. When this sector display is used, a triple intersection, which represents a target, will move vertically on the display when the range VFO is tuned and horizontally when the velocity VFO is adjusted, indicating the movement of the "window" over the FM/CW plane.

Fig. 7 is a block diagram of a possible receiver system which uses sector coverage.

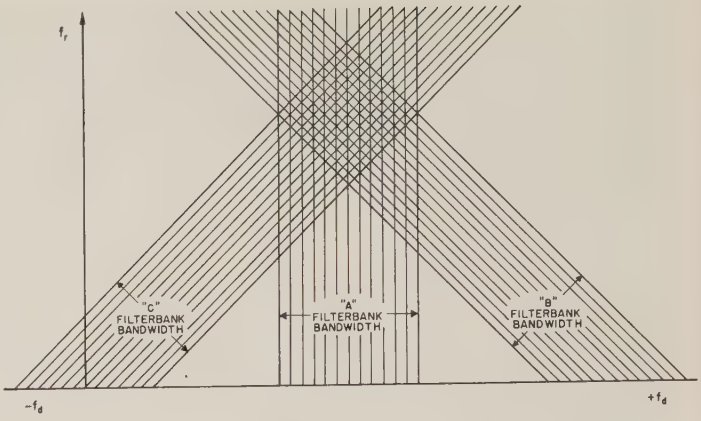


Fig. 5—Sector coverage in FM/CW plane.

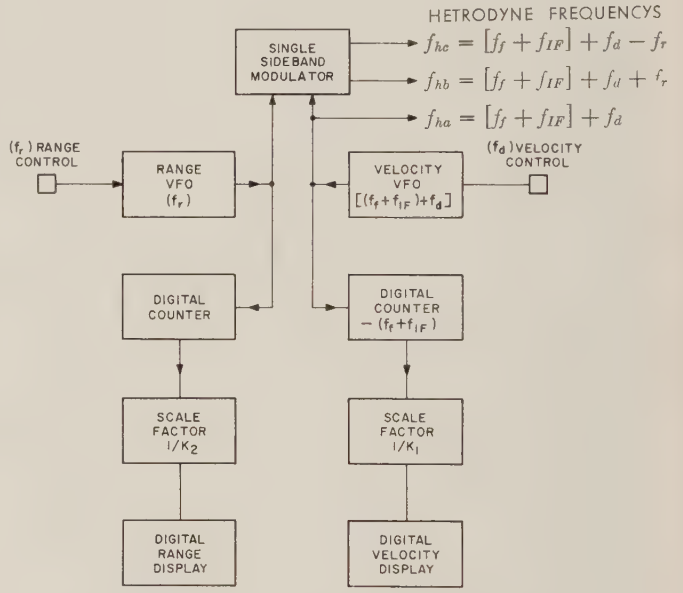


Fig. 6—VFO control for sector coverage.

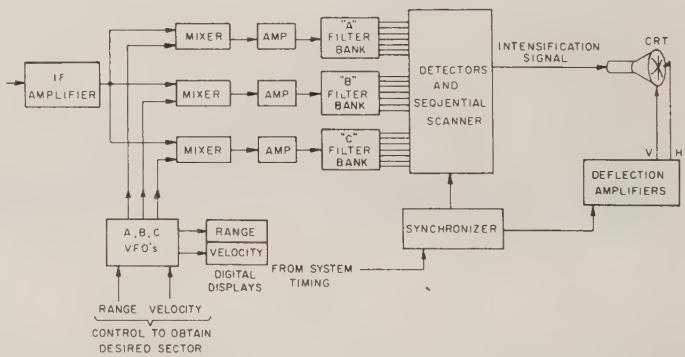


Fig. 7—Block diagram—an FM/CW system with sector coverage.

VII. Maximum Utilization of Available Filters

One interesting aspect of sector coverage is the determination of the maximum utilization of a given number of filters. Fig. 5 illustrates the filter coverage when an equal number of filters are used in the A, B, and C filter banks. It is apparent that the filters at the high-

and low-frequency end of the A filter bank contribute little to the area in which it is possible to obtain a triple intersection. If these filters were removed and transferred to the B and C banks, this area is increased and the "window" shape becomes hexagonal.

With reference to Fig. 8, the following derivation indicates that for a given number of filters, the maximum FM/CW coverage is obtained when the B and C banks each contain 50 per cent more filters than the A bank.

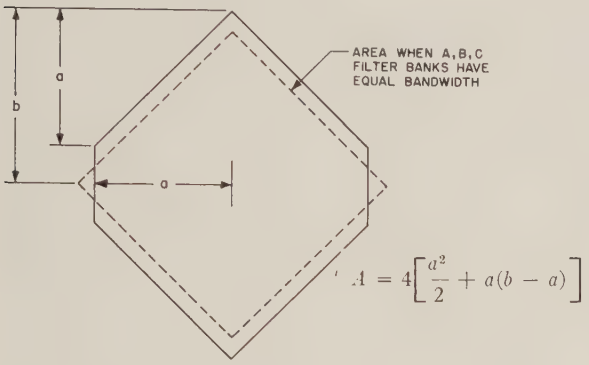


Fig. 8—Maximum filter utilization in FM/CW system.

If A is the area of the hexagon, $2a$ is the number of filters in the A bank, $2b$ is the number of filters in the B or C bank and N is the total number of available filters.

$$N = 2a + 4b \quad \text{and} \quad b = \frac{N - 2a}{4} \tag{7}$$

The area $A = 4[\frac{1}{2}b^2 + a(b-a)]$. Simplifying and substituting for b,

$$A = -(4a^2 + aN), \tag{8}$$

after differentiating and setting equal to zero,

$$0 = \frac{dA}{da} = -8a + N \quad \text{and} \quad \frac{b}{a} = 1.5. \tag{9}$$

VIII. Expanded Tracking Displays

A useful tracking or magnified display can be obtained by modifying the timing of the synchronizing triggers. With reference to Fig. 3, a marker generator can be built which will allow superposition of the marker and the target returns. The marker generator has two delay controls. When the A mode is being scanned, the marker signal appears at a time t_d after the start of A-filter scan and the B marker appears at a time $(t_a + t_r)$ after the start of the B-filter scan. Similarly, the marked delay after the start of C scan is equal to $(t_d - t_r)$. The target can be tracked by means of the range and velocity delay controls. The delays t_d, t_r can be measured, and by suitable proportionality constants, the range and velocity of the tracked target can be displayed or recorded. These numbers can be added to those obtained from the VFO's for absolute determination of the target's velocity and range.

In order to increase the tracking accuracy and obtain a magnified 10× display, a pretrigger is generated by the marker generator 1/20 of the time it takes to scan a filter bank. If these A, B and C pretriggers initiate a sweep on a panoramic display similar to Fig. 3 with a duration 1/10 of the filter scan time, the markers always will appear in the center of this expanded display; and when marker tracking is on target, the A, B and C returns will be centered and positioned one above the other.

If this sweep voltage, which is synchronized to the pretriggers, is used in place of the horizontal deflection voltage used in the FM/CW display, an expanded display results covering 1/100 the area of the FM/CW plane display. Adjustment of the marker controls t_d, t_r will position this smaller portion of the FM/CW plane within that portion covered by the filter banks in Fig. 5. This positioning is similar to the orthogonal "window" positioning by VFO control and is illustrated in Fig. 5. When the markers are positioned to coincide with the returns from a particular target, the triple intersection will appear at the exact center of this magnified tracking FM/CW display. Several of these tracking displays can be operated simultaneously with the addition of a marker generator and display for each tracking operator.

IX. Results—Implementation

The FM/CW plane display was built using a twelve-inch, electrostatically deflected CRT, and is illustrated in Figs. 4, 9 and 10. In Fig. 9, a simulator is used to generate eleven signals in the A, B and C filter banks, dividing each area into ten equal spaces. The resultant test pattern illustrates the ability of the display to accurately resolve triple intersections over the face of the tube. In the design of such an indicator there are two basic types of distortion to consider. One type is conveniently called "rubber" distortion, for it is equivalent to having the FM/CW plane graph drawn on a sheet of rubber. Stretching this sheet can cause various types of inaccuracies but will not destroy a triple intersection. This type of distortion occurs when the nonlinearities are introduced in stages that are commonly shared by the A, B and C deflection signals. The geometric distortions in the CRT, and nonlinearities in the deflection-driver amplifiers are examples of "rubber" distortion. An examination of Fig. 9 indicates the presence of "barrel" type distortion which does not deteriorate the triple intersections. When distortions are introduced in deflection signal paths which are not common to all three modes, it is possible that an exact point of triple will not occur. Care was taken to reduce this form of distortion, and the results indicated in Fig. 9 were obtained with comparatively simple circuitry with good reliability and a minimum of adjustments.

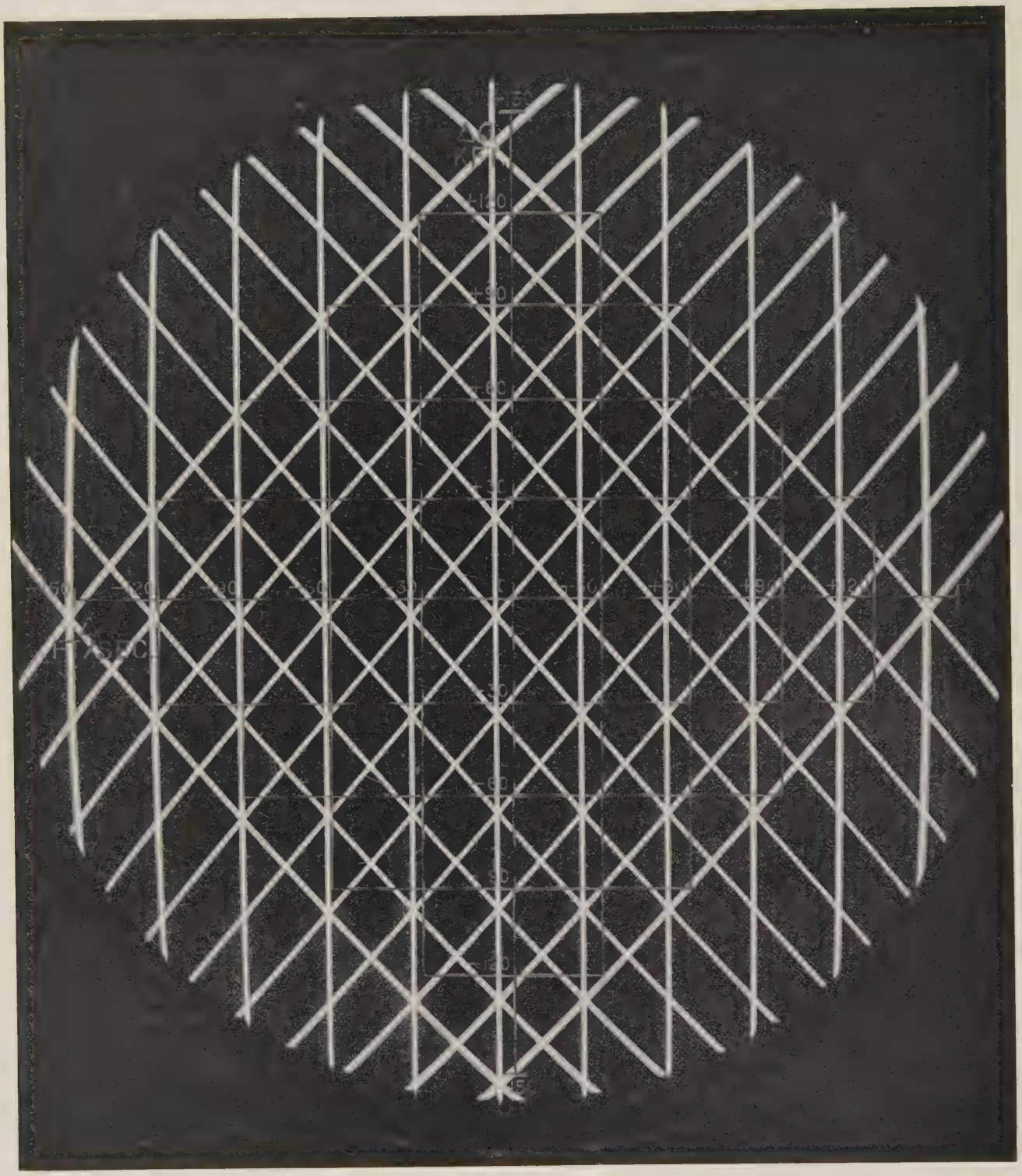


Fig. 9—Triple-intersection test pattern.

Fig. 10 is a photograph of the results obtained during a controlled two-target flight test. The differences in range and velocity between these two targets are clearly evident. The panoramic display at the right was energized with the identical filter output signals used for intensification of the FM/CW display. Examination of the panoramic display illustrates the difficulty an operator would have in utilizing the spectrum information under multiple-target conditions. The sharp marker "pips" controlled by t_d, t_r are also evident on this display.

X. Future Developments

As the target's coordinate information is contained only in the point of triple intersection, it would be desirable to eliminate all the excess lines from the display to reduce the confusion when there are a large number of targets. Several systems for accomplishing this end

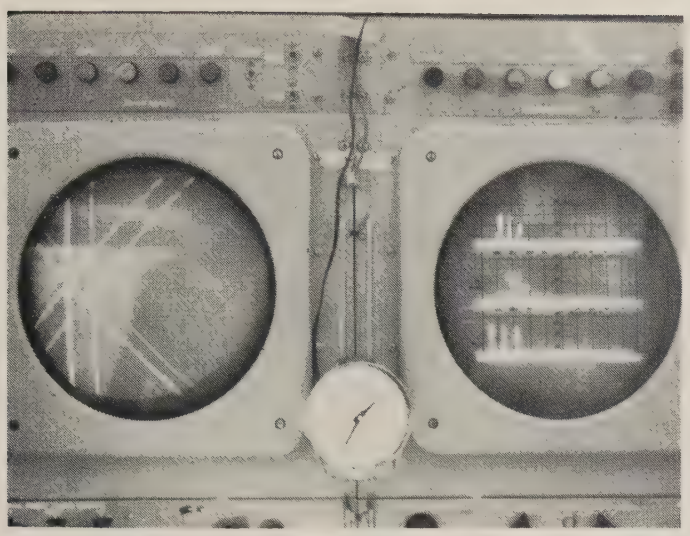


Fig. 10—Multiple-target flight test.

have been considered. One system in particular is being presently pursued. In addition to eliminating the excess lines, it has the capabilities of automatic readout of multiple target coordinates and adjustable persistence enabling the target coordinate point to be displayed as a "track."

Other systems than those indicated in Fig. 7 are applicable. A ramp with twice the frequency deviation of the B and C modes can be substituted for the CW mode. Returns would be plotted as a line with twice the slope of those in the B and C modes.

Acknowledgment

The author wishes to thank Col. Nicholson and Capt. George of RADC under whose cognizance some of these developments were accomplished, L. Pugliese, S. Eisenmesser, J. Sur, A. Plevy for their contributions to this program, and W. Heiser and R. Bailey for their advice and encouragement.

Bibliography

[1] L. N. Ridenour, "Radar System Engineering," MIT Rad. Lab. Ser., Cambridge, Mass.; 1948.
[2] D. G. Luck, "Frequency Modulated Radar," McGraw-Hill Book Co., Inc., New York, N. Y.; 1949.
[3] R. I. Bernstein, Univ. of Mich., 4th Annual Radar Symposium, Ann Arbor, Mich.; June, 1958.
[4] H. H. Naidich, IT & T Internal. Memo-66872, IT & T, Nutley, N. J ; December, 1958.

Contributors

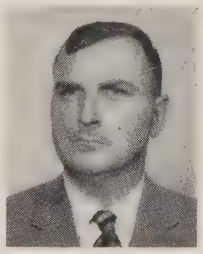
Robert Adler (A'42–SM'47–F'51) was born in Vienna, Austria, on December 4, 1913. He received the Ph.D. degree in physics from the University of Vienna in 1937.

R. Adler

During the following year he was assistant to a patent attorney in Vienna. From 1939 to 1940 he worked in the laboratory of Scientific Acoustic, Ltd., London, England. After one year with Associated Research, Inc., Chicago, Ill., he joined the research group of Zenith Radio Corporation, Chicago, in 1941; he became Zenith's Associate Director of Research in 1952. His previously published work in the vacuum-tube field includes the development of the phasitron modulator, of receiving tubes such as the 6BN6 and 6AR8, and of transverse-field traveling-wave tubes.

Dr. Adler is a member of the AAAS.

John D. Albright (S'47–A'48–M'52) was born in Richmond, Va., on March 2, 1923. He received the B.S.E.E. and M.S.E.E. degrees from the Georgia Institute of Technology, Atlanta, in 1944 and 1947.



J. D. Albright

From 1944–1945 he was with the Western Electric Company, Kearny, N. J. From 1947–1954 he was with the NOL, White Oak, Md., the NRL and the Diamond Ordnance Fuze Laboratory, Washington, D. C., working mainly on the design of projectile and missile fuzes. Since 1954 he has been with the Westingthouse Electric Corporation, Air Arm Division, Baltimore, Md., as a Senior Engineer engaged in various aspects of the design of airborne fire control radar systems. He has devoted the past two years to the special problems of pulse-Doppler systems.

Mr. Albright is a member of the AIEE.

Lamont V. Blake was born in Somerville, Mass., on Nov. 7, 1913. He received the B.S. degree from Massachusetts State College (now the University of Massachusetts), Amherst, in 1935, and the M.S. degree from the University of Maryland, College Park, in 1950, both in physics.

L. V. Blake

From 1937 to 1940 he was employed by the Arkansas Power and Light Company as a radio interference investigator. Since 1940 he has been at the Naval Research Laboratory in Washington, D. C., where he is now a Consultant in the Search Radar Branch, Radar Diyision, and where he has worked on problems of radar transmitter design, radar countermeasure techniques, radar system design and evaluation, and the theory of radar detection. Part-time activities include teaching for the University of Maryland in 1952 and 1953 and writing

lesson-text material for the Capitol Radio Engineering Institute in 1960.

Mr. Blake is a Registered Professional Engineer in the District of Columbia. He is a member of the American Physical Society, AAAS, RESA, Sigma Pi Sigma, and the Philosophical Society of Washington.

❖

Charles R. Boyd, Jr., (M'56) was born in Pittsburgh, Pa., on October 21, 1932. He received the B.S.E.E. degree from Carnegie Institute of Technology, Pittsburgh, Pa., in 1953. During his undergraduate studies, he held Carnegie Freshman, Buhl Foundation, Pennsylvania State, and U. S. Steel Corporation scholarships. In 1959 he completed the three-year General Electric Company Advanced Engineering Program, and he is currently engaged in a program of graduate studies at Syracuse University, Syracuse, N. Y.

C. R. Boyd, Jr.

From 1953 until 1956 he was employed by the Westinghouse Electric Corporation, Baltimore, Md., where he worked as a field engineer on developmental autopilot and radar systems. In 1956 he joined the General Electric Light Military Electronics Department, Utica, N. Y., as a design engineer on the Atlas ICBM Guidance Program. In 1957 he transferred to his present position on the technical staff of the General Electric Electronics Laboratory, Syracuse, N. Y., where he is engaged in the study and development of microwave systems and networks. His most recent work has been concerned with the development of parametric amplifiers, harmonic generators, and high-speed microwave switches.

Mr. Boyd is a Licensed Professional Engineer in New York State, and is a member of Eta Kappa Nu.

❖

John S. Burgess (SM'60) was born in Milwaukee, Wis., on May 1, 1918. He received the B.S. degree in mathematics from St. Lawrence University, Canton, N. Y., in 1940, the M. S. degree in physics from the University of Notre Dame, Notre Dame, Ind., in 1942, and the Ph.D. degree in physics from The Ohio State University, Columbus, in 1949.

J. S. Burgess

From 1940 to February, 1951, he was employed for four years in industry as a research assistant for the General Electric Company, in Schenectady, N. Y., and during the remainder of the time was engaged in research and teaching at the University of Notre Dame, The Ohio State University, and St. Lawrence University.

He has been in governmental service since February, 1951, and presently is assigned to the Rome Air Development Center, Griffiss Air Force Base, N. Y., as Scientific Director. In this capacity he advises the Commander on all technical matters, and reviews and approves broad objectives for the development program of the Center.

Dr. Burgess is a member of the American Institute of Physics, Phi Beta Kappa, Sigma Xi, Sigma Pi Sigma and Pi Mu Epsilon.

❖

Kenneth W. Cowans was born in New York City, N. Y., on December 27, 1932. He received the B.S. degree in mechanical engineering with high honors from North Carolina State College, Raleigh, in 1957.

K. W. Cowans

He joined Hughes Aircraft Company, Culver City, Calif., in 1957, where he has been engaged in research on several cryogenic projects.

Mr. Cowans is a member of Tau Beta Pi, Pi Tau Sigma, and Phi Kappa Phi.

❖

Louis J. Cutrona (M'46–SM'47) was born in Buffalo, N. Y., on March 11, 1915. He received the B.A. degree in physics from Cornell University, Ithaca, N. Y., in 1936, and the M.A. and Ph.D. degrees in physics from the University of Illinois, Urbana, in 1938 and 1940, respectively.

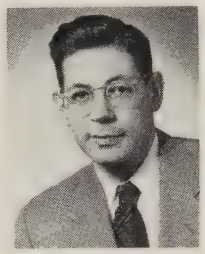
L. J. Cutrona

From 1940 to 1942 he was an Instructor of physics and electrical engineering at the Duluth Junior College, Duluth, Minn. From 1942 to 1943 he was a member of the staff of the York Safe and Lock Company, York, Pa., and from 1943 to 1945 he was a member of the technical staff of Bell Telephone Laboratories, New York, N. Y. From 1945 to 1947 he was an employee of the Federal Telecommunications Corporation, Nutley, N. J.; from 1947 to 1949 of the Sperry Gyroscope Company, Great Neck, N. Y.; and in 1949 he joined the Willow Run Laboratories of The University of Michigan, Ann Arbor. He has since been associated with The University of Michigan, except for the year 1956–1957, when he was on leave of absence with the Space Technology Laboratory, Los Angeles, Calif. In 1957 he received an appointment as Professor of electrical engineering at The University of Michigan. For the year 1959–1960 he was on leave from that position as Head of the Radar Laboratory of The University of Michigan, Willow Run Laboratories. His work has been in a variety of fields, particularly those of circuit theory, information theory, communications, radar, and missile system design.

Dr. Cutrona is a member of Phi Kappa Phi, Pi Mu Epsilon, Sigma Xi, Eta Kappa Nu, Tau Beta Pi, The American Physical Society, and the American Institute of Physics.

❖

Charles A. Fowler (A'46–M'55) was born in Centralia, Ill., on December 17, 1920. He received the B.S. degree in engineering physics from the University of Illinois, Urbana, in 1942.



C. A. Fowler

From 1942 to 1945 he was at the Radiation Laboratory, M.I.T., Cambridge, Mass. where he participated in the development of the GCA (ground-controlled approach) radar system and was project engineer on the development of the AN/CPN-18(XW-1) air traffic control radar. Since 1946 he has been with the Airborne Instruments Laboratory, Melville, N. Y., where he has worked on air traffic control studies, air defense problems, and radar techniques and systems, and where he is presently Head of the Radar Systems Department.

❖

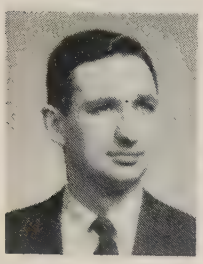
Ernst Gehrels was born in San Francisco, Calif., on May 16, 1930. He received the B.S. degree in physics from the California Institute of Technology, Pasadena, Calif., in 1952, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif., in 1958.

E. Gehrels

While at Stanford Dr. Gehrels worked on a project involving meteor propagation of radio signals. He also participated in research on magnetic-ionic ducting effects at VLF, spending part of the International Geophysical Year in Tierra del Fuego, S.A., in connection with this work. In 1958 he joined the Surveillance Techniques Group of the Lincoln Laboratory, M.I.T., Lexington, Mass. where he is active in the radio physics and radar astronomy program of the Millstone Hill Radar Station.

Dr. Gehrels is a member of Sigma Xi.

Norman R. Gillespie (A'54–SM'59) was born in Youngstown, Ohio, on February 7, 1929. He received the B.S.E.E. and M.S., degrees from Northwestern University, Evanston, Ill., in 1953 and 1960, respectively.



N. R. Gillespie

From 1953 to 1960 he was employed by the Cook Research Laboratories, Morton Grove, Ill., where he shared responsibility for the evaluation and reduction of counter-measures susceptibility in radar systems, and where he also directed development of target and ECM simulators for operator training. He is presently a systems project engineer in the Ordnance Radar Department of the Raytheon Company, Wayland, Mass.

❖

Louis P. Goetz (SM'52–M'46) was born in Shelley, Ida., on November 9, 1915. He received the B.S.E.E. degree from the University of Idaho, Moscow, in 1938, and the M.S.E.E. degree from Oklahoma State University, Stillwater, in 1950.

L. P. Goetz

In 1939 he was employed by Utah Power and Light Company, Salt Lake City, and in 1940 he joined the Westinghouse Electric Corporation, East Pittsburgh, Pa., where he was engaged in engineering tests. From 1941 to 1945 he served in the Signal Corps of the U. S. Army. During most of this period he was at Evans Signal Laboratory, Belmar, N. J., taking part in radar systems development. After World War II, until 1949, he performed instrument design work in the Research Department of Phillips Petroleum Company, Bartlesville, Okla. From 1951 to 1955 he worked in the Research Department of Stanolind Oil and Gas Company, Tulsa, Okla., engaged in analysis and instrument design. In 1955 he joined the Air Arm Division of Westinghouse Electric Corporation in the Radar Development Section, Baltimore, Md., where he is a staff engineer engaged in pulse-Doppler radar systems design.

Mr. Goetz is a member of the American Society of Professional Engineers.

❖

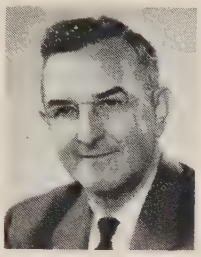
Frank E. Goodwin (S'57–M'58) was born in Hastings, Neb., on September 7, 1927. He received the M.S. degree in applied physics from the University of California at Los Angeles in 1957.

He joined the Hughes Research Laboratories, Malibu, Calif., in 1957. During 1957 and 1958 he conducted applied research on electronic scanning of microwave antennas. Since 1959 he has been Head of the Maser Applications Group, concerned with the development of solid-state masers and their applications to radar and allied systems.

F. E. Goodwin

❖

Gilbert O. Hall (A'47–M'55) was born in Mattawan, Mich., on October 31, 1907. He received the B.S.E.E. degree from Michigan State College, East Lansing, in 1929, and the M.S. degree in engineering sciences from Harvard University, Cambridge, Mass., in 1948.



G. O. Hall

From 1930 to 1936 he was with General Electric Company, first as a test engineer and later as a field product representative. He served as a Naval Officer in World War II, teaching in the radar schools at Harvard and M.I.T., where he was a contributing author for the Radiation Laboratory series. From 1946 to 1951 he worked for the Air Force Cambridge Research Laboratories in the field of upper atmosphere research with rockets. In 1951 he joined the staff of the Willow Run Laboratories of the University of Michigan, Ann Arbor, where, as a research engineer, he has been concerned with research in the fields of radar and radar countermeasures, and where he is also a Lecturer in the Electrical Engineering Department.

Mr. Hall is a member of Tau Beta Pi and Sigma Xi.

❖

John B. Higley was born in Kennewick Wash., on February 5, 1911. He received the B.S.E.E. degree from the California Institute of Technology, Pasadena, in 1935.

J. B. Higley

From 1935 to 1941 he worked with the U. S. Reclamation Bureau as a civil engineer on construction of the All-American Canal in Imperial Valley, Calif. While in the U. S. Army from 1941 to 1946, his chief assignment was at the Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, as Project Officer with the Radar Trainer Group. He has been project engineer and Assistant Head of the R & D Department of the American Machine and Foundry Company, Boston, Mass., where he developed radar trainers and video-mapping equipment. He was engaged in sonar design and development at Massa Laboratories, Hingham, Mass., as Chief Electrical Engineer. Since 1954 he has been a staff engineer with the Raytheon Company, Wayland, Mass., where he is engaged in the design and development of fire-control radar and weapon-control systems.

❖

E. J. Kelly was born in Philadelphia, Pa., on November 4, 1924. He received the B. S. and Ph.D. degrees in physics from the Massachusetts Institute of Technology, Cambridge, in 1945 and 1950, respectively.

E. J. Kelly

From 1950 to 1952 he was employed at the Brookhaven National Laboratory, Brookhaven, N. Y. Since 1952 he has been with the Lincoln Laboratory, M.I.T., Lexington, Mass., where he is at present Associate Leader of the Radar Receivers Group.

❖

James E. Kiefer was born in Wichita, Kan., on August 27, 1931. He received the B.S. degree in engineering physics from the University of Tulsa, Tulsa, Okla., in 1958, and the M.S. degree in applied physics from the University of California at Los Angeles in 1960.

J. E. Kiefer

In 1958 he joined the Hughes Aircraft Company, Culver City, Calif., as a participant in their Master of Science Fellowship Program, and is now a member of the technical staff of the Hughes Research Laboratories, Malibu, Calif. Since 1958 he has been engaged in the development of solid-state masers and their applications to microwave systems, and has also been concerned with microwave, magnetic and cryogenic techniques as applied to masers.

Mr. Kiefer is a member of the American Physical Society.

Aharon A. Ksienski (S'53–M'58) was born in Warsaw, Poland, on June 24, 1924. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1952 and 1958, respectively.

A. A. Ksienski

From 1948 to 1951 he was in charge of an Aircraft Electrical School in the Israeli Air Force. From 1951 to 1952 he attended the University of Southern California as a graduate student and part-time member of the teaching staff. From 1953 to 1955 he was with Wiancko Engineering Company, Pasadena, Calif., engaged in the development of instrumentation systems for the measurement of various flight parameters and in the design of electromechanical transducers. He returned to the University of Southern California in 1955 as a Lecturer in mathematics and electrical engineering, and continued graduate research in the field of antennas. In 1958 he joined the Antenna Research Department in the Microwave Laboratory of Hughes Aircraft Company, Culver City, Calif., where he is presently a Senior Staff Engineer engaged in research in the field of antennas and antenna systems.

Dr. Ksienski is a member of Phi Kappa Phi, Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

❖

Emmett N. Leith was born in Detroit, Mich., on March 12, 1927. He received the B.S. and M.S. degrees from Wayne State University, Detroit, Mich., in 1949 and 1952, respectively.

E. N. Leith

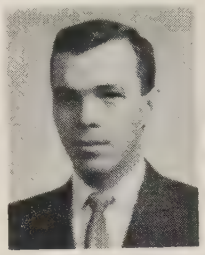
In 1952 he joined the staff of the Willow Run Laboratories at The University of Michigan, Ann Arbor, where he is presently employed. His work has been in the fields of radar, communications, data processing and optics.

Mr. Leith is a member of Sigma Xi.

❖

Norman MacKinnon was born in Wilmington, Del., on July 24, 1931. He received the B.S.E.E. degree from Rutgers University, New Brunswick, N. J., in 1953.

He then worked at the Hazeltine Electronics Division of the Hazeltine Corporation, Little Neck, N. Y., as a design engineer on the design, development and field evaluation of airborne ASW and IFF systems. In 1956 he joined the staff of the Raytheon Company, Wayland, Mass., where his duties have included project engineering responsibilities on missile guidance radar system design, and where he is presently head of a systems project engineering section in the Ordnance Radar Department.



N. MacKinnon

❖

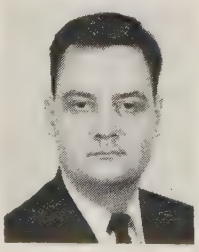
Herbert H. Naidich was born in New York, N. Y., on May 6, 1924. He received the B.E.E. degree from the College of the City of New York in 1949, and the M.E.E. from the Newark College of Engineering, Newark, N. J.

H. Naidich

From 1942 to 1946 he served in the Army, working on communications and fire control equipment. After graduation, he joined the Civil Aeronautics Authority, engaging in radar development at the General Electric Company, Electronics Park, Syracuse, N. Y. From 1950 to 1955 he was with the Bendix Radio Division, Towson, Md., where he was active in the development of several radar systems. Since 1955 he has been with the International Telephone and Telegraph Labs., Nutley, N. J., engaged in the development of several advanced countermeasures and radar systems.

❖

Frank C. Ogg, Jr., (M'56–SM'59) was born in Champaign, Ill., on January 22, 1930. He received the B.S. in mathematics from Bowling Green State University, Bowling Green, Ohio, in 1951. From 1951 to 1955 he was Junior Instructor in mathematics at The Johns Hopkins University, Baltimore, Md., where he received the Ph.D. degree in mathematics in 1955.



F. C. Ogg, Jr.

From 1955 to 1960 he was employed by the Bendix Corporation, Baltimore, Md., as a research engineer and later as Principal Research Engineer. In this capacity he worked on radar data processing, pulse compression, noise analysis and antijamming circuitry for radar. From 1958 to 1959 he was Supervisor of the Systems Group on a large steerable array radar project. In 1960 he was appointed Research Scientist in the Radiation Laboratory of The Johns Hopkins University, where he is now Supervisor of the Signal Analysis Group.

Dr. Ogg is a member of the American Mathematical Society, the Society for Industrial and Applied Mathematics, and Sigma Xi.

❖

Austin Parsons (M'57) was born in Hartford, Conn., on April 10, 1930. He received the B.A. degree in physics from Boston University, Boston, Mass., in 1956.

A. Parsons

From 1956 to 1957 he was employed at the Ewen Knight Corporation, Needham, Mass., and from 1957 to 1960 he was a staff member of the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington. He is currently again employed at the Ewen Knight Corporation.

❖

Herbert O. Ramp (M'59) was born in Vienna, Austria, on August 9, 1924. He received the Diplom Ingenieur degree from the University of Technology of Vienna, Austria, in 1951.

H. O. Ramp

After graduation he joined the staff of the Receiving Tubes Laboratory of Siemens and Halske, Vienna, where he was in charge of the Measurement Group. In 1953 he was employed by the Signal Corps Engineering Laboratories at Fort Monmouth, N. J., where he was project engineer for passive radar development and battlefield surveillance radar. In 1956 he joined the Electronics Laboratory of the General Electric Company, Syracuse, N. Y., where he has been engaged in developmental work in the pulse compression field since 1957.

❖

Gerald F. Ross (S'51–A'53–M'57–SM'60) was born in New York, N. Y., on December 14, 1930. He received the B.E.E. degree from the College of the City of New York in 1952,

and the M.E.E. degree from the Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in 1955, where he is currently a doctoral candidate.

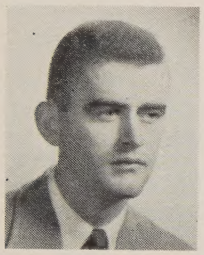
G. F. Ross

In 1952 he was employed as a Research Assistant at the University of Michigan Laboratories, Ann Arbor, where he developed microwave receivers. In 1953 he served briefly as an electronics project officer at the U. S. Air Force Holloman Air Development Center, Alamogordo, N. Mex., and later that same year joined the W. L. Maxson Corporation, New York, N. Y., where he helped develop the ALQ-23 ECM system. In 1958 he joined the Sperry Gyroscope Company, Great Neck, N. Y., where he continued his work in the ECM field and where he has recently been engaged in advanced radar studies.

Mr. Ross is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

❖

John A. Ruetz (S'50–A'50–M'56) was born in Remus, Mich., on August 21, 1926. He received the B.S.E.E. and M.S.E.E. degrees from the University of Michigan, Ann Arbor, in 1949 and 1950, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif., in 1957.

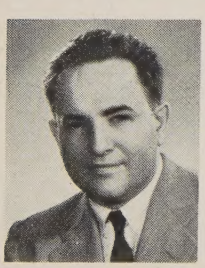


J. A. Ruetz

From 1950 to 1953 he worked on low-noise traveling-wave tubes at RCA Laboratories, Princeton, N. J. From 1953 to 1957 he was employed in the Electronics Research Laboratory at Stanford University. Since 1957 he has been employed by Varian Associates, Palo Alto, Calif., where he is now working on the development of high-power traveling-wave tubes.

❖

Abraham E. Ruvin (A'46–M'55) was born in Brooklyn, N. Y., on May 17, 1921. He received the B.E.E. degree from the Cooper Union, New York, N. Y., in 1942, and the M.E.E. degree from the Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in 1949.



A. E. Ruvin

From 1942 to 1946 he was an engineer with the National Advisory Committee for Aeronautics at Langley Field, Virginia, where he worked on the development of a radio telemetering system used for measurements in free flight and rocket missiles. From 1947 to 1950 he was an engineer at Hazeltine Electronics Corporation, Little Neck, N. Y., where he participated in development of IFF systems and radar relay receivers for an AEW system. Since 1950 he has been with the Airborne Instruments Laboratory, Melville, N. Y., where he has been a project engineer in the development of advanced MTI techniques and missile defense study, and where, at present, he is a section head in the Radar Systems Department.

❖

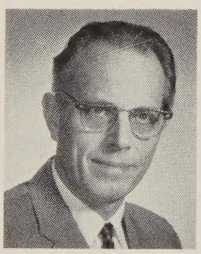
Leon Schwartzman (S'56–M'59) was born in Brooklyn, N. Y., on February 6, 1931. He received the B.E.E. degree from the Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in 1958.

L. Schwartzman

From 1952 to 1954 he served in the U. S. Navy as a radarman. Since 1957 he has been with the Microwave Electronics Department of the Sperry Gyroscope Company, Great Neck, N. Y., where he has worked on the development of microwave antennas, specializing in monopulse systems, cassegrainian antennas, and wide-angle scanning systems. He is presently engaged in electromagnetic wave propagation, ionospheric physics, and communication problems.

❖

Henry R. Senf (A'38–SM'46) was born in Louisville, Ky., on July 3, 1912. He received the B.S. degree in physics from Antioch College, Yellow Springs, Ohio, in 1936.



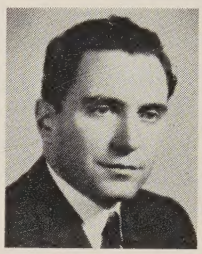
H. R. Senf

From 1934 to 1938 he assisted in the development of instrument landing systems in the Aircraft Radio Laboratory, Wright Field, Ohio. From 1938 to 1947 he was employed at the Naval Research Laboratory, Washington, D. C., where he worked on radar altimeters, airborne radar, and the Mark V IFF system. From 1947 to 1949 he developed high-speed computing circuitry for the SEAC computer at the National Bureau of Standards, Washington, D. C. During the next five years he was a member of the technical staff of the Air Navigation Development Board, where he was concerned with numerous air navigation and traffic control projects. In 1954 he joined Lockheed Missile Systems Division, Van Nuys, Calif., as Head of the Electronics Department.

Since 1956 he has been associated with the Research Laboratories of Hughes Aircraft Company, Malibu, Calif., where he has been concerned primarily with applied research on electronic antenna scanning, three-dimensional cathode-ray tube displays, and solid-state maser amplifiers.

❖

Murray Simpson (M'46) was born in New York, N. Y., on July 27, 1921. He received the B.E.E. degree from the College of the City of New York in 1942, and the M.E.E. degree from the Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in 1952.



M. Simpson

For a number of years he has been concerned with the design of advanced radar systems and electronic countermeasures equipment. Prior to his entrance into the Navy, he was with the Federal Telecommunications Laboratories, New York, N. Y., where he was concerned with the development of various types of aircraft navigation equipments, such as the omni-range and instrument landing system. As a naval lieutenant during World War II he was assigned to the Aircraft Radio Section of the Naval Research Laboratory, Washington, D. C., where he developed various types of airborne antennas and microwave-filter networks. Immediately after the war, he took part in the early development of microwave communications systems at Raytheon Manufacturing Company, Waltham, Mass. From 1948–1950 he headed a group at Fairchild Guided Missiles Company concerned with the development of a surface-to-air missile system. At present, he is Assistant Vice-President and Technical Director of Maxson Electronics Corporation, New York, N. Y., responsible for the technical performance of all research and development programs.

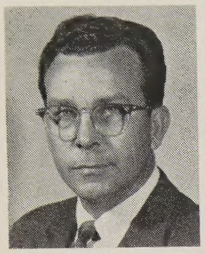
Mr. Simpson is a member of Tau Beta Pi, Eta Kappa Nu, and the American Ordnance Association.

❖

Austin R. Sisson (S'47–A'50–SM'55) was born in Harveyton, Ky., on February 24, 1923. He received the B.S.E.E. and M.S.E.E. degrees from the University of Tennessee, Knoxville, in 1948 and 1949, respectively.

He was a research engineer at the University of Tennessee from 1949 to 1950. During 1950 and 1951 he was employed by Douglas Aircraft Company, El Segundo, Calif., as a design engineer in the antenna group. From 1951 to 1953 he was a Senior

Engineer with Microwave Engineering Company, Los Angeles, Calif., where he designed microwave systems and antennas.



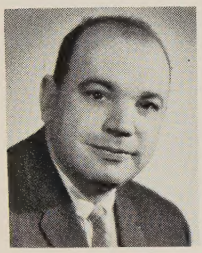
A. R. Sisson

In 1953 he joined the Bendix-Pacific Division of the Bendix Corporation, North Hollywood, Calif., where he is now Head of the Microwave Department, responsible for microwave systems and antenna engineering in the areas of radar, missile guidance, beacons, altimeters, and telemetry.

Mr. Sisson is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.

❖

Anthony P. Uzzo, Jr., (A'50–M'58) was born in Wilmington, Del., on May 11, 1923. He received the B.S. degree in engineering from the University of Delaware, Newark, in 1948.



A. P. Uzzo, Jr.

While at the University of Delaware, he participated in the development of instrumentation for flame study research. Following graduation, he joined the Airborne Instruments Laboratory, Melville, N. Y., where he has worked on MTI, surveillance radar systems, and automatic tracking devices, and where he has been project engineer on MTI developments and on the design of the radar processing equipment for the USAF Frequency Diversity Program. Recent efforts include the direction of production of the ASDE II and AN/FPN-31 radars, the AN/PPS-5 radar, and the AN/GPA-98 simulation system. At present, he is a section head in the Department of Radar Systems.

❖

William S. Van Slyck (SM'56) was born in Ironwood, Mich., on September 22, 1921. He received the B.S.E.E. degree from the University of Michigan, Ann Arbor, in 1943. He then attended Harvard University and the Massachusetts Institute of Technology, Cambridge, for special radar, sonar, and communications study.

W. S. Van Slyck

While in the Navy, he served as radar officer in submarine service. Since 1947 he has been employed by Zenith Radio Corporation, Chicago, Ill., as a development engineer, and is now Chief Electrical Engineer of the Military Division.

Mr. Van Slyck is a member of the Armed Forces Communications and Electronics Association.

❖

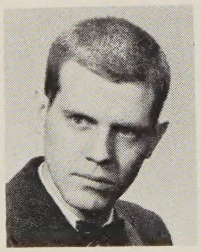
Weston E. Vivian was born in Newfoundland, Can., on October 25, 1924. He received the B.S.E.E. degree from Union College, Schenectady, N. Y., in 1945, the S.M. degree from Massachusetts Institute of Technology, Cambridge, in 1949, and the Ph.D. degree from The University of Michigan, Ann Arbor, in 1959.

W. E. Vivian

After release from service as an officer in the U. S. Navy in 1946, he was engaged in research in electronics for the Sperry Gyroscope Co., Great Neck, N. Y., Massachusetts Institute of Technology, Boeing Airplane Co., Seattle, Wash., and The University of Michigan, successively through 1951. He is presently Vice President for Engineering, Conductron Corp., Ann Arbor, Mich. His technical activities have included development of guidance and reconnaissance systems, and research on microwave properties of ionized gases and beam devices.

❖

E. R. Wingrove, Jr., (S'51–M'56) was born in McDonald, Pa., on February 24, 1931. He received the B.S. degree in 1952, the M.S. degree in 1953, and the Ph.D. degree in 1955, all from the Carnegie Institute of Technology, Pittsburgh, Pa. From 1952 to 1954, he was a Research Assistant, and from 1954 to 1955, a Buhl Fellow, at Carnegie, where he worked on problems of instrumentation and measurement of magnetic and dielectric amplifiers.



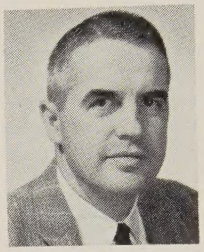
E. Wingrove, Jr.

Since 1955 he has been a member of the staff of the General Electric Electronics Laboratory, Syracuse, N. Y., where he is presently a project engineer engaged in advanced development of pulse compression radar techniques, monopulse radar systems, three-coordinate radar, limiting amplifiers, phase detectors, and precision phase measurement techniques.

Dr. Wingrove is a member of Sigma Xi and RESA.

❖

W. H. Yocom was born in Oberlin, Ohio, on May 15, 1919. He received the B.A. degree in physics from Oberlin College, in 1940, the B.S.E.E. degree from the Massachusetts Institute of Technology, Cambridge, in 1942, and the M.S.E.E. degree from Stevens Institute of Technology, Hoboken, N. J., in 1950.



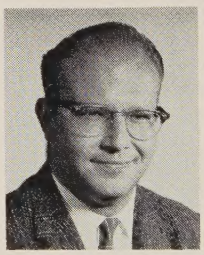
W. H. Yocom

From 1942 to 1956 Mr. Yocom was employed by Bell Telephone Laboratories, Murray Hill, N. J., where he was first concerned with carrier and microwave systems, and later with microwave tube research, particularly beam focusing and generation of millimeter waves. In 1956 he joined Varian Associates, Palo Alto, Calif., where he currently heads the Wave Tube Research and Development Department.

Mr. Yocom is a member of Sigma Xi.

❖

Gregory O. Young (S'43–A'47–SM'53) was born in Spokane, Wash., on April 20, 1922. He received the B.S.E.E. and M.S.E.E. degrees from the California Institute of Technology, Pasadena, in 1944 and 1947, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1956.



G. O. Young

From 1947 to 1956 he was a member of the technical staff of the Hughes Aircraft Company, Culver City, Calif., and a Lecturer at U.S.C. Since 1956 he has been an Associate Professor of electrical engineering at U.S.C. and a Consultant at Hughes. His principal interests and experience lie in the fields of information theory, noise, systems analysis and servo design. He is presently engaged in research on data-processing antennas and information theory applied to antenna design.

Dr. Young is a member of Tau Beta Pi, Sigma Xi, and Eta Kappa Nu.

INFORMATION FOR AUTHORS

The PGMIL TRANSACTIONS is intended to bridge the gap between the various disciplines contributing to military electronics. Since this includes most of the branches of electronics, of the military services, and of the many fields which are associated with but not actually within the realm of electronics, it is essential that the papers published be of broad interest. The emphasis should be on readable, thought-provoking material that stimulates an attitude of open-mindedness and curiosity.

The major portion of the PGMIL TRANSACTIONS publication program is in the form of special issues designed to bring together the technical achievements of one field or by one group of workers. Topics and guest editors for forthcoming issues are announced regularly nine months in advance of the publication date. Each issue is open to contributions from anyone working in the area covered by that issue. Detailed abstracts of all contributed papers must be submitted for review at least eight weeks prior to the manuscript deadline. Abstracts and manuscripts should be sent in duplicate directly to the guest editor in charge of the issue of interest. Standard IRE practice should be followed in preparation of the manuscript and illustrations. Each manuscript should include a carefully prepared summary of not more than 200 words.

Suggestions are earnestly solicited from the membership on topics for future issues of the PGMIL TRANSACTIONS. Suggested topics should be sent to the PGMIL TRANSACTIONS Editor, Donald R. Rhodes, Radiation Incorporated, P.O. Box 6904, Orlando, Fla.

PUBLICATION SCHEDULE

Publication Date	*Topic*	*Guest Editor*	*Manuscript Deadline*
October, 1961	"Missile and Space Range Instrumentation" Radar, optics, telemetry, timing, intra range communications, frequency coordination, trajectory computation, etc.	Mr. A. G. Waggoner Asst. Director (Ranges and Space Ground Support) Office of Director of Defense Research and Engineering Washington 25, D. C.	July 1, 1961
January, 1962	"Direct Energy Conversion" Processes and devices: photoelectricity, thermoelectricity, thermionics, fuel cells, galvanic batteries, magneto hydrodynamics.	Mr. G. B. Wareham Office of Fuels, Materials, and Ordnance Office of Director of Defense Research and Engineering Washington 25, D. C.	October, 1 1961
	Next Issue		
July, 1961	"Micro Electronics and Systems" Solid-state circuit techniques and their influence on future system design.	Dr. J. E. Thomas, Jr. Director of Research and Engrg. Semiconductor Division Sylvania Electric Products, Inc. Woburn, Mass.	

INSTITUTIONAL LISTINGS

The IRE Professional Group on Military Electronics is grateful for the assistance given by the firms listed below, and invites application for Institutional Listings from other firms interested in the field of Military Electronics.

PHILCO CORP., Government and Industrial Div., 4700 Wissahickon Ave., Philadelphia 44, Pa.
Microwave, Radar, Computers, Guided Missiles, Space Instrumentation, and Other Military Electronics Systems

THE RAMO-WOOLDRIDGE LABORATORIES, 8433 Fallbrook Ave., Canoga Park, Calif.
Electronic Systems Technology and Intellectronics Laboratories

TEXAS INSTRUMENTS INCORPORATED, Apparatus Div., 600 Lemmon Ave., Dallas 9, Tex.
Missile Electronics, Submarine Detection Systems, Electronic Surveillance Systems, Heavy Radar

The charge for an Institutional Listing is $75.00 per issue or $225.00 for four consecutive issues. Applications for Institutional Listings and checks (made out to the Institute of Radio Engineers) should be sent to Mr. L. G. Cumming, Professional Groups Secretary, Institute of Radio Engineers, 1 East 79 Street, New York 21, N. Y.